

1-1-2022

Interacting with educational chatbots: A systematic review

Mohammad Amin Kuhail
Zayed University

Nazik Alturki
Princess Nourah bint Abdulrahman University

Salwa Alramlawi
Princess Nourah bint Abdulrahman University

Kholood Alhejori
Independent Scholar

Follow this and additional works at: <https://zuscholars.zu.ac.ae/works>



Part of the [Computer Sciences Commons](#)

Recommended Citation

Kuhail, Mohammad Amin; Alturki, Nazik; Alramlawi, Salwa; and Alhejori, Kholood, "Interacting with educational chatbots: A systematic review" (2022). *All Works*. 5208.
<https://zuscholars.zu.ac.ae/works/5208>

This Article is brought to you for free and open access by ZU Scholars. It has been accepted for inclusion in All Works by an authorized administrator of ZU Scholars. For more information, please contact scholars@zu.ac.ae.



Interacting with educational chatbots: A systematic review

Mohammad Amin Kuhail¹ · Nazik Alturki² · Salwa Alramlawi³ · Kholood Alhejori⁴

Received: 28 February 2022 / Accepted: 14 June 2022
© The Author(s) 2022

Abstract

Chatbots hold the promise of revolutionizing education by engaging learners, personalizing learning activities, supporting educators, and developing deep insight into learners' behavior. However, there is a lack of studies that analyze the recent evidence-based chatbot-learner interaction design techniques applied in education. This study presents a systematic review of 36 papers to understand, compare, and reflect on recent attempts to utilize chatbots in education using seven dimensions: educational field, platform, design principles, the role of chatbots, interaction styles, evidence, and limitations. The results show that the chatbots were mainly designed on a web platform to teach computer science, language, general education, and a few other fields such as engineering and mathematics. Further, more than half of the chatbots were used as teaching agents, while more than a third were peer agents. Most of the chatbots used a predetermined conversational path, and more than a quarter utilized a personalized learning approach that catered to students' learning needs, while other chatbots used experiential and collaborative learning besides other design principles. Moreover, more than a third of the chatbots were evaluated with experiments, and the results primarily point to improved learning and subjective satisfaction. Challenges and limitations include inadequate or insufficient dataset training and a lack of reliance on usability heuristics. Future studies should explore the effect of chatbot personality and localization on subjective satisfaction and learning effectiveness.

Keywords Chatbot · Conversational Agent · Educational Bot · Literature Review · Interaction Styles · Human-Computer Interaction

✉ Nazik Alturki
namalturki@pnu.edu.sa

Extended author information available on the last page of the article

1 Introduction

Chatbots, also known as conversational agents, enable the interaction of humans with computers through natural language, by applying the technology of natural language processing (NLP) (Bradeško & Mladenović, 2012). Due to their ability to emulate human conversations and thus automate services and reduce effort, chatbots are increasingly becoming popular in several domains, including healthcare (Oh et al., 2017), consumer services (Xu et al., 2017), education (Angheliescu & Nicolaescu, 2018), and academic advising (Alkhoori et al., 2020). In fact, the size of the chatbot market worldwide is expected to be 1.23 billion dollars in 2025 (Kaczorowska-Spychalska, 2019). In the US alone, the chatbot industry was valued at 113 million US dollars and is expected to reach 994.5 million US dollars in 2024 ¹.

The adoption of educational chatbots is on the rise due to their ability to provide a cost-effective method to engage students and provide a personalized learning experience (Benotti et al., 2018). Chatbot adoption is especially crucial in online classes that include many students where individual support from educators to students is challenging (Winkler & Söllner, 2018). Chatbots can facilitate learning within the educational context, for instance by instantaneously providing students with course content (Cunningham-Nelson et al., 2019), assignments (Ismail & Ade-Ibijola, 2019), rehearsal questions (Sinha et al., 2020), and study resources (Mabunda, 2020). Moreover, chatbots may interact with students individually (Hobert & Meyer von Wolff, 2019) or support collaborative learning activities (Chaudhuri et al., 2009; Tegos et al., 2014; Kumar & Rose, 2010; Stahl, 2006; Walker et al., 2011). Chatbot interaction is achieved by applying text, speech, graphics, haptics, gestures, and other modes of communication to assist learners in performing educational tasks.

Existing literature review studies attempted to summarize current efforts to apply chatbot technology in education. For example, Winkler and Söllner (2018) focused on chatbots used for improving learning outcomes. On the other hand, Cunningham-Nelson et al. (2019) discussed how chatbots could be applied to enhance the student's learning experience. The study by Pérez et al. (2020) reviewed the existing types of educational chatbots and the learning results expected from them. Smutny and Schreiberova (2020) examined chatbots as a learning aid for Facebook Messenger. Thomas (2020) discussed the benefits of educational chatbots for learners and educators, showing that the chatbots are successful educational tools, and their benefits outweigh the shortcomings and offer a more effective educational experience. Okonkwo and Ade-Ibijola (2021) analyzed the main benefits and challenges of implementing chatbots in an educational setting.

The existing review studies contributed to the literature, albeit their main emphasis was using chatbots for improving the learning experience and outcomes (Winkler & Söllner, 2018; Cunningham-Nelson et al., 2019; Smutny & Schreiberova, 2020; Thomas, 2020), identifying the types of educational chatbots (Pérez et al., 2020), and determining the benefits, and challenges of implementing educational chatbots

¹ <https://www.statista.com/statistics/656596/worldwidechatbot-market/>

(Okonkwo & Ade-Ibijola, 2021). Nonetheless, the existing review studies have not concentrated on the chatbot interaction type and style, the principles used to design the chatbots, and the evidence for using chatbots in an educational setting.

Given the magnitude of research on educational chatbots, there is a need for a systematic literature review that sheds light on several vital dimensions: field of application, platform, role in education, interaction style, design principles, empirical evidence, and limitations.

By systematically analyzing 36 articles presenting educational chatbots representing various interaction styles and design approaches, this study contributes: (1) an in-depth analysis of the learner-chatbot interaction approaches and styles currently used to improve the learning process, (2) a characterization of the design principles used for the development of educational chatbots, (3) an in-depth explanation of the empirical evidence used to back up the validity of the chatbots, and (4) the discussion of current challenges and future research directions specific to educational chatbots. This study will help the education and human-computer interaction community aiming at designing and evaluating educational chatbots. Potential future chatbots might adopt some ideas from the chatbots surveyed in this study while addressing the discussed challenges and considering the suggested future research directions. This study is structured as follows: In Section 2, we present background information about chatbots, while Section 3 discusses the related work. Section 4 explains the applied methodology, while Section 5 presents the study's findings. Section 6 presents the discussion and future research directions. Finally, we present the conclusion and the study's limitations in Section 7.

2 Background

Chatbots have existed for more than half a century. Prominent examples include ELIZA, ALICE, and SmarterChild. ELIZA, the first chatbot, was developed by Weizenbaum (1966). The chatbot used pattern matching to emulate a psychotherapist conversing with a human patient. ALICE was a chatbot developed in the mid-1990s. It used Artificial Intelligence Markup Language (AIML) to identify an accurate response to user input using knowledge records (AbuShawar and Atwell, 2015). Another example is Smart Child (Chukhno et al., 2019), which preceded today's modern virtual chatbot-based assistants such as Alexa² and Siri³, which are available on messaging applications with the ability to emulate conversations with quick data access to services.

Chatbots have been utilized in education as conversational pedagogical agents since the early 1970s (Laurillard, 2013). Pedagogical agents, also known as intelligent tutoring systems, are virtual characters that guide users in learning environments (Seel, 2011). Conversational Pedagogical Agents (CPA) are a subgroup of pedagogical agents. They are characterized by engaging learners in a dialog-based

² <https://developer.amazon.com/en-US/alexa>

³ <https://www.apple.com/sa-ar/siri/>

conversation using AI (Gulz et al., 2011). The design of CPAs must consider social, emotional, cognitive, and pedagogical aspects (Gulz et al., 2011; King, 2002).

A conversational agent can hold a discussion with students in a variety of ways, ranging from spoken (Wik & Hjalmarsson, 2009) to text-based (Chaudhuri et al., 2009) to nonverbal (Wik & Hjalmarsson, 2009; Ruttkay & Pelachaud, 2006). Similarly, the agent's visual appearance can be human-like or cartoonish, static or animated, two-dimensional or three-dimensional (Dehn & Van Mulken, 2000). Conversational agents have been developed over the last decade to serve a variety of pedagogical roles, such as tutors, coaches, and learning companions (Haake & Gulz, 2009). Furthermore, conversational agents have been used to meet a variety of educational needs such as question-answering (Feng et al., 2006), tutoring (Heffernan & Croteau, 2004; VanLehn et al., 2007), and language learning (Heffernan & Croteau, 2004; VanLehn et al., 2007).

When interacting with students, chatbots have taken various roles such as teaching agents, peer agents, teachable agents, and motivational agents (Chhibber & Law, 2019; Baylor, 2011; Kerry et al., 2008). *Teaching agents* play the role of human teachers and can present instructions, illustrate examples, ask questions (Wambsganss et al., 2020), and provide immediate feedback (Kulik & Fletcher, 2016). On the other hand, *peer agents* serve as learning mates for students to encourage peer-to-peer interactions. The agent of this approach is less knowledgeable than the teaching agent. Nevertheless, peer agents can still guide the students along a learning path. Students typically initiate the conversation with peer agents to look up certain definitions or ask for an explanation of a specific topic. Peer agents can also scaffold an educational conversation with other human peers.

Students can teach *teachable agents* to facilitate gradual learning. In this approach, the agent acts as a novice and asks students to guide them along a learning route. Rather than directly contributing to the learning process, motivational agents serve as companions to students and encourage positive behavior and learning (Baylor, 2011). An agent could serve as a teaching or peer agent and a motivational one.

Concerning their interaction style, the conversation with chatbots can be chatbot or user-driven (Følstad et al., 2018). Chatbot-driven conversations are scripted and best represented as linear flows with a limited number of branches that rely upon acceptable user answers (Budi, 2018). Such chatbots are typically programmed with if-else rules. When the user provides answers compatible with the flow, the interaction feels smooth. However, problems occur when users deviate from the scripted flow.

User-driven conversations are powered by AI and thus allow for a flexible dialogue as the user chooses the types of questions they ask and thus can deviate from the chatbot's script. There are one-way and two-way user-driven chatbots. One-way user-driven chatbots use machine learning to understand what the user is saying (Dutta, 2017), and the responses are selected from a set of premade answers. In contrast, two-way user-driven chatbots build accurate answers word by word to users (Winkler & Söllner, 2018). Such chatbots can learn from previous user input in similar contexts (De Angeli & Brahnam, 2008).

In terms of the medium of interaction, chatbots can be text-based, voice-based, and embodied. Text-based agents allow users to interact by simply typing via a

Table 1 Areas relevant studies focused on

Area of coverage	Relevant Studies
Field of application	(Winkler & Söllner, 2018; Wollny et al., 2021; Hwang & Chang, 2021; Smutny & Schreiberova, 2020)
Objectives and learning experiences	(Winkler & Söllner, 2018; Cunningham-Nelson et al., 2019; Pérez et al., 2020; Wollny et al., 2021; Hwang & Chang, 2021)
How it is applied	(Winkler & Söllner, 2018; Cunningham-Nelson et al., 2019; Wollny et al., 2021)
Design approaches	(Winkler & Söllner, 2018; Martha & Santoso, 2019; Hwang & Chang, 2021)
Technology used	(Pérez et al., 2020)
Evaluation methods	(Pérez et al., 2020; Hwang & Chang, 2021; Hobert & Meyer von Wolff, 2019)
Challenges	(Okonkwo & Ade-Ibijola, 2021)

keyboard, whereas voice-based agents allow talking via a mic. Voice-based chatbots are more accessible to older adults and some special-need people (Brewer et al., 2018). An embodied chatbot has a physical body, usually in the form of a human, or a cartoon animal (Serenko et al., 2007), allowing them to exhibit facial expressions and emotions.

Concerning the platform, chatbots can be deployed via messaging apps such as Telegram, Facebook Messenger, and Slack (Car et al., 2020), standalone web or phone applications, or integrated into smart devices such as television sets.

3 Related work

Recently several studies reviewed chatbots in education. The studies examined various areas of interest concerning educational chatbots, such as the field of application (Smutny & Schreiberova, 2020; Wollny et al., 2021; Hwang & Chang, 2021), objectives and learning experience (Winkler & Söllner, 2018; Cunningham-Nelson et al., 2019; Pérez et al., 2020; Wollny et al., 2021; Hwang & Chang, 2021), how chatbots are applied (Winkler & Söllner, 2018; Cunningham-Nelson et al., 2019; Wollny et al., 2021), design approaches (Winkler & Söllner, 2018; Martha & Santoso, 2019; Hwang & Chang, 2021), the technology used (Pérez et al., 2020), evaluation methods used (Pérez et al., 2020; Hwang & Chang, 2021; Hobert & Meyer von Wolff, 2019), and challenges in using educational chatbots (Okonkwo & Ade-Ibijola, 2021). Table 1 summarizes the areas that the studies explored.

Winkler and Söllner (2018) reviewed 80 articles to analyze recent trends in educational chatbots. The authors found that chatbots are used for health and well-being advocacy, language learning, and self-advocacy. Chatbots are either flow-based or powered by AI, concerning approaches to their designs.

Several studies have found that educational chatbots improve students' learning experience. For instance, Okonkwo and Ade-Ibijola (2021) found out that chatbots motivate students, keep them engaged, and grant them immediate assistance,

particularly online. Additionally, Wollny et al. (2021) argued that educational chatbots make education more available and easily accessible.

Concerning how they are applied, Cunningham-Nelson et al. (2019) identified two main applications: answering frequently-asked questions (FAQ) and performing short quizzes, while Wollny et al. (2021) listed three other applications, including scaffolding, activity recommendations, and informing them about activities.

In terms of the design of educational chatbots, Martha and Santoso (2019) found out that the role and appearance of the chatbot are crucial elements in designing the educational chatbots, while Winkler and Söllner (2018) identified various types of approaches to designing educational chatbots such as flow and AI-based, in addition to chatbots with speech recognition capabilities.

Pérez et al. (2020) identified various technologies used to implement chatbots such as Dialogflow⁴, FreeLing (Padró and Stanilovsky, 2012), and ChatFuel⁵. The study investigated the effect of the technologies used on performance and quality of chatbots.

Hobert and Meyer von Wolff (2019), Pérez et al. (2020), and Hwang and Chang (2021) examined the evaluation methods used to assess the effectiveness of educational chatbots. The authors identified that several evaluation methods such as surveys, experiments, and evaluation studies measure acceptance, motivation, and usability.

Okonkwo and Ade-Ibijola (2021) discussed challenges and limitations of chatbots including ethical, programming, and maintenance issues.

Although these review studies have contributed to the literature, they primarily focused on chatbots as a learning aid and thus how they can be used to improve educational objectives. Table 2 compares this study and the related studies in terms of the seven dimensions that this study focuses on: field of application, platform, educational role, interaction style, design principles, evaluation, and limitations.

Only four studies (Hwang & Chang, 2021; Wollny et al., 2021; Smutny & Schreiberova, 2020; Winkler & Söllner, 2018) examined the field of application. None of the studies discussed the platforms on which the chatbots run, while only one study (Wollny et al., 2021) analyzed the educational roles the chatbots are playing. The study used “teaching,” “assisting,” and “mentoring” as categories for educational roles. This study, however, uses different classifications (e.g., “teaching agent”, “peer agent”, “motivational agent”) supported by the literature in Chhibber and Law (2019), Baylor (2011), and Kerlyl et al. (2006). Other studies such as (Okonkwo and Ade-Ibijola, 2021; Pérez et al., 2020) partially covered this dimension by mentioning that chatbots can be teaching or service-oriented.

Only two articles partially addressed the interaction styles of chatbots. For instance, Winkler and Söllner (2018) classified the chatbots as flow or AI-based, while Cunningham-Nelson et al. (2019) categorized the chatbots as machine-learning-based or dataset-based. In this study, we carefully look at the interaction style in

⁴ <https://dialogflow.cloud.google.com>

⁵ <https://chatfuel.com/>

Table 2 Comparison between this work and relevant studies

Reference	Field of Application	Platform	Educational Role	Interaction Style	Design Principles	Evaluation	Limitations
(Winkler & Söllner, 2018)	✓	-	-	Partial	Partial	-	-
(Hobert & Meyer von Wolff, 2019)	-	-	-	-	-	Partial	-
(Martha & Santos, 2019)	-	-	-	-	Partial	-	-
(Cunningham-Nelson et al., 2019)	-	-	-	Partial	-	-	-
(Pérez et al., 2020)	-	-	Partial	-	-	✓	✓
(Smutny & Schreiberova, 2020)	✓	-	Partial	-	-	✓	-
(Wollny et al., 2021)	✓	-	✓	-	Partial	-	-
(Okonkwo & Ade-Ibijola, 2021)	-	-	Partial	-	-	-	Partial
(Hwang & Chang, 2021)	✓	-	-	-	-	Partial	-
This study	✓	✓	✓	✓	✓	✓	✓

terms of who is in control of the conversation, i.e., the chatbot or the user. As such, we classify the interactions as either chatbot or user-driven.

Only a few studies partially tackled the principles guiding the design of the chatbots. For instance, Martha and Santoso (2019) discussed one aspect of the design (the chatbot's visual appearance). This study focuses on the conceptual principles that led to the chatbot's design.

In terms of the evaluation methods used to establish the validity of the articles, two related studies (Pérez et al., 2020; Smutny & Schreiberova, 2020) discussed the evaluation methods in some detail. However, this study contributes more comprehensive evaluation details such as the number of participants, statistical values, findings, etc.

Regarding limitations, Pérez et al. (2020) examined the technological limitations that have an effect on the quality of the educational chatbots, while Okonkwo and Ade-Ibijola (2021) presented some challenges and limitations facing educational chatbots such as ethical, technical, and maintenance matters. While the identified limitations are relevant, this study identifies limitations from other perspectives such as the design of the chatbots and the student experience with the educational chatbots. To sum up, Table 2 shows some gaps that this study aims at bridging to reflect on educational chatbots in the literature.

4 Methodology

The literature related to chatbots in education was analyzed, providing a background for new approaches and methods, and identifying directions for further research. This study follows the guidelines described by Keele et al. (2007). The process includes these main steps: (1) defining the review protocol, including the research questions, how to answer them, search strategy, and inclusion and exclusion criteria. (2) running the study by selecting the articles, assessing their quality, and synthesizing the results. (3) reporting the findings.

4.1 Research questions

Based on the shortcomings of the existing related literature review studies, we formulated seven main research questions:

RQ1 - In what fields are the educational chatbots used?

RQ2 - What platforms do the educational chatbots operate on?

RQ3 - What role do the educational chatbots play when interacting with students?

RQ4 - What are the interaction styles supported by the educational chatbots?

Table 3 Inclusion and Exclusion criteria

Inclusion Criteria (IC)	Exclusion Criteria (EC)
IC-1: The article is written in English	EC-1: The duplicate studies with the same content.
IC-2: The article introduces an educational chatbot.	EC-2: The article is a tutorial, poster, technical report, review paper, short paper, or a Ph.D. thesis.
IC-3: The article thoroughly explains the details of the usage of the educational chatbot.	EC-3: Duplicate article that presents a chatbot introduced in another article (only the most recent article is included in this case)
	EC-4: The article presented an educational chatbot but with no or little empirical study to back its validity.)

RQ5 - What are the principles used to guide the design of the educational chatbots?

RQ6 - What empirical evidence exists to support the validity of the educational chatbots?

RQ7 - What are the challenges of applying and using the chatbots in the classroom?

The first question identifies the fields of the proposed educational chatbots, while the second question presents the platforms the chatbots operate on, such as web or phone-based platforms. The third question discusses the roles chatbots play when interacting with students. For instance, chatbots could be used as teaching or peer agents. The fourth question sheds light on the interaction styles used in the chatbots, such as flow-based or AI-powered. The fifth question addresses the principles used to design the proposed chatbots. Examples of such principles could be collaborative and personalized learning. The sixth question focuses on the evaluation methods used to prove the effectiveness of the proposed chatbots. Finally, the seventh question discusses the challenges and limitations of the works behind the proposed chatbots and potential solutions to such challenges.

4.2 Search process

The search process was conducted during the period (2011 - 2021) in the following databases: ACM Digital Library, Scopus, IEEE Xplore, and SpringerLink. We analyzed our research questions, objectives, and related existing literature review studies to identify keywords for the search string of this study. Subsequently, we executed and refined the keywords and the search string iteratively until we arrived at promising results. We used these search keywords: "Chatbot" and "Education." Correlated keywords for "Chatbot" are "Conversational Agent" and "Pedagogical Agent." Further, correlated keywords for "Education" are "Learning," "Learner," "Teaching," "Teacher," and "Student."

The search string was defined using the Boolean operators as follows: ('Chatbot' OR 'Conversational Agent' OR 'Pedagogical Agent') AND ('Education' OR 'Learning' OR 'Learner' OR 'Teaching' OR 'Teacher' OR 'Student')

According to their relevance to our research questions, we evaluated the found articles using the inclusion and exclusion criteria provided in Table 3. The inclusion and exclusion criteria allowed us to reduce the number of articles unrelated to our research questions. Further, we excluded tutorials, technical reports, posters, and Ph.D. thesis since they are not peer-reviewed.

After defining the criteria, our search query was performed in the selected databases to begin the inclusion and exclusion process. Initially, the total of studies resulting from the databases was 1208 studies. The metadata of the studies containing; title, abstract, type of article (conference, journal, short paper), language, and keywords were extracted in a file format (e.g., bib file format). Subsequently, it was imported into the Rayyan tool ⁶, which allowed for reviewing, including, excluding, and filtering the articles collaboratively by the authors.

The four authors were involved in the process of selecting the articles. To maintain consistency amongst our decisions and inter-rater reliability, the authors worked in two pairs allowing each author to cross-check the selection and elimination of the author they were paired with. The process of selecting the articles was carried out in these stages:

1. Reading the articles' metadata and applying the inclusion criteria of IC-1 and the exclusion criteria of EC-1. As a result, the number of studies was reduced to 1101.
2. As a first-round, we applied the inclusion criterion IC-2 by reading the studies' title, abstract, and keywords. Additionally, the EC-2 exclusion criterion was applied in the same stage. As a result, only 197 studies remained.
3. In this stage, we eliminated the articles that were not relevant to any of our research questions and applied the EC-3 criteria. As a result, the articles were reduced to 71 papers.
4. Finally, we carefully read the entire content of the articles having in mind IC-3. Additionally, we excluded studies that had no or little empirical evidence for their effectiveness of the educational chatbot (EC-4 criterion). As a result, the articles were reduced to 36 papers.

Figure 1. shows the flowchart of the selecting processes, in which the final stage of the selection resulted in 36 papers.

⁶ <https://www.rayyan.ai/>

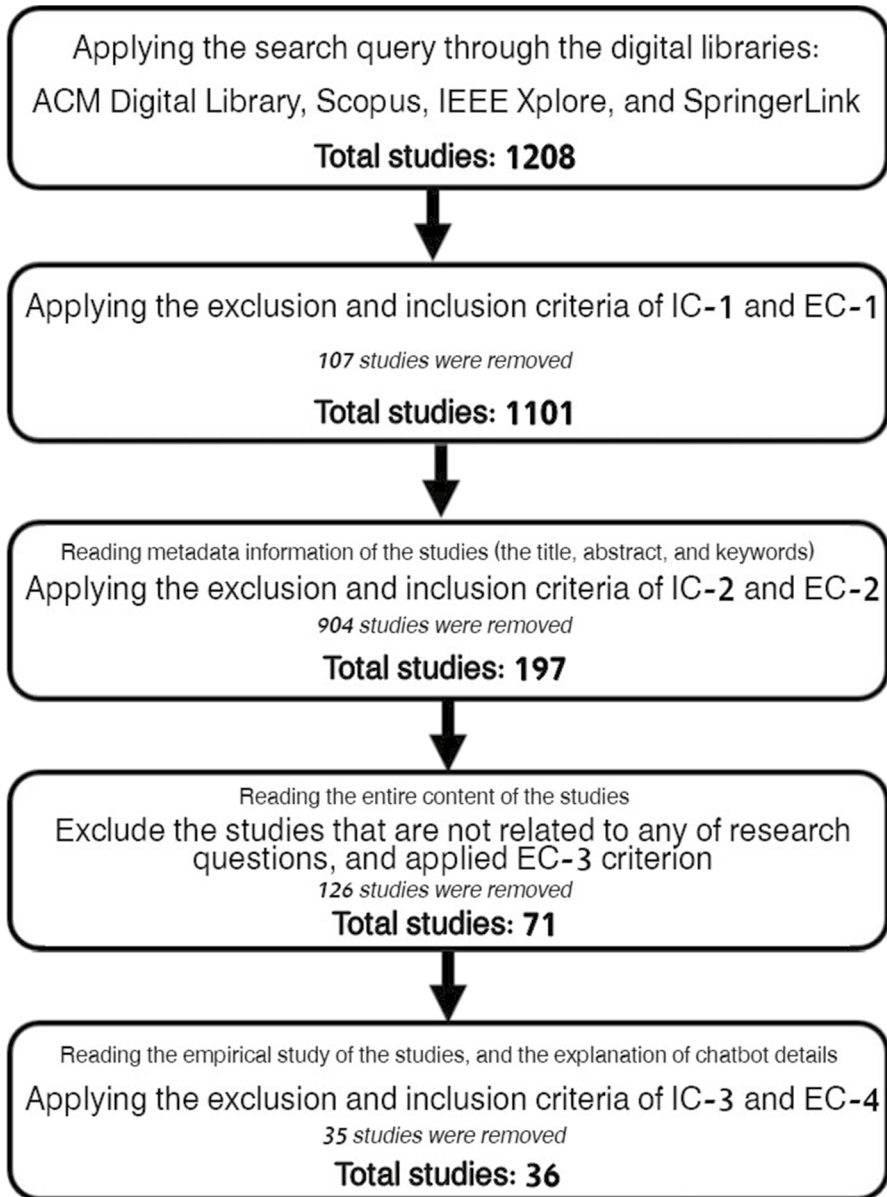


Fig. 1 Flowchart of the process of the selection of the studies

5 Results

Figure 2 shows the number and types of articles plotted against time. 63.88% (23) of the selected articles are conference papers, while 36.11% (13) were published in journals. Most conference papers were published after 2017. Interestingly, 38.46%

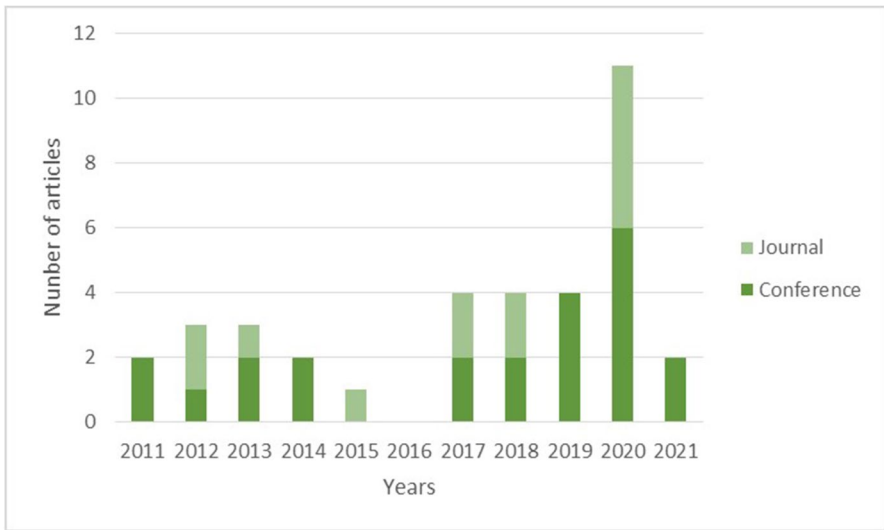


Fig. 2 A timeline of the selected studies

(5) of the journal articles were published recently in 2020. Concerning the publication venues, two journal articles were published in the Journal of IEEE Transactions on Learning Technologies (TLT), which covers various topics such as innovative online learning systems, intelligent tutors, educational software applications and games, and simulation systems for education. Intriguingly, one article was published in Computers in Human Behavior journal. The remaining journal articles were published in several venues such as IEEE Transactions on Affective Computing, Journal of Educational Psychology, International Journal of Human-Computer Studies, ACM Transactions on Interactive Intelligent System. Most of these journals are ranked Q1 or Q2 according to Scimago Journal and Country Rank ⁷.

Figure 3. shows the geographical mapping of the selected articles. The total sum of the articles per country in Fig. 3 is more than 36 (the number of selected articles) as the authors of a single article could work in institutions located in different countries. The vast majority of selected articles were written or co-written by researchers from American universities. However, the research that emerged from all European universities combined was the highest in the number of articles (19 articles). Asian universities have contributed 10 articles, while American universities contributed 9 articles. Further, South American universities have published 5 articles. Finally, universities from Africa and Australia contributed 4 articles (2 articles each).

⁷ <https://www.scimagojr.com/>

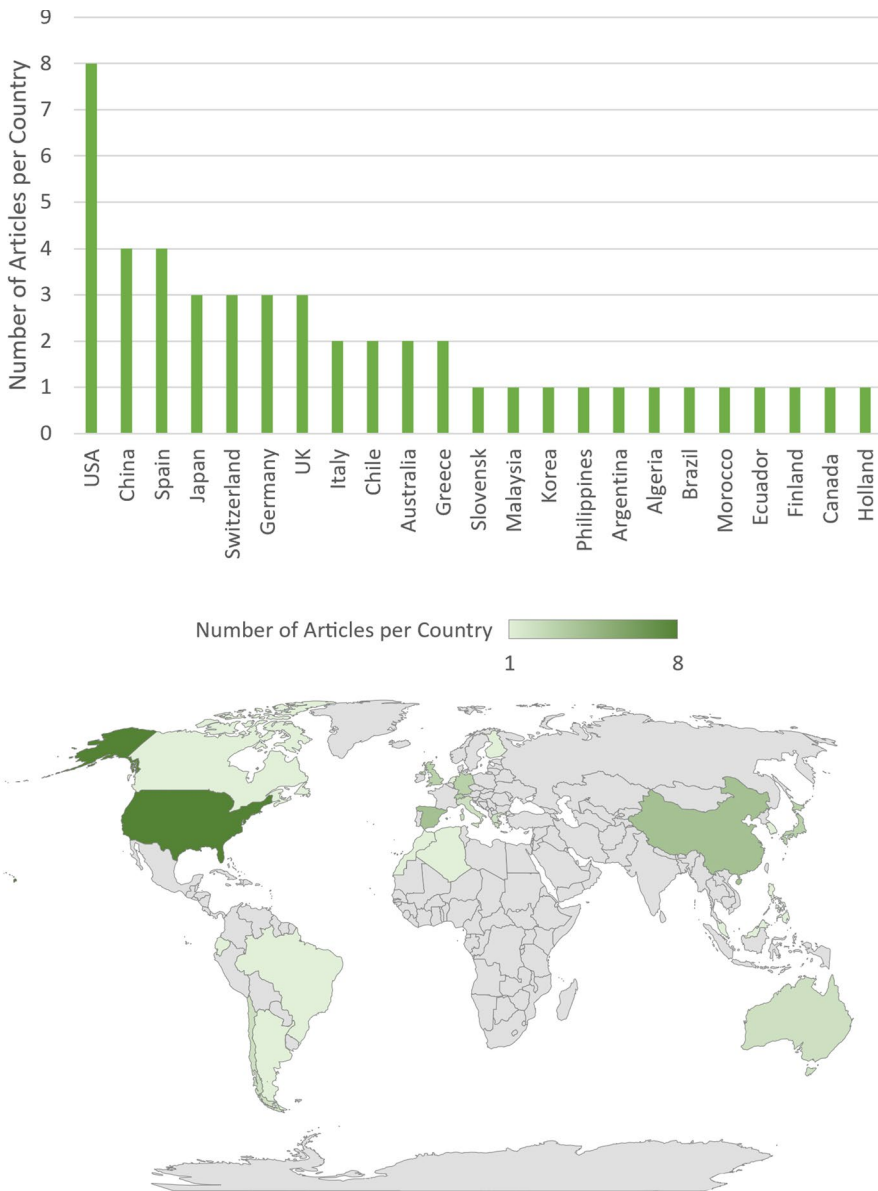


Fig. 3 A geographical mapping of the selected articles

5.1 RQ1: What fields are the educational chatbots used in?

Recently, chatbots have been utilized in various fields (Ramesh et al., 2017). Most importantly, chatbots played a critical role in the education field, in which most researchers (12 articles; 33.33%) developed chatbots used to teach computer science

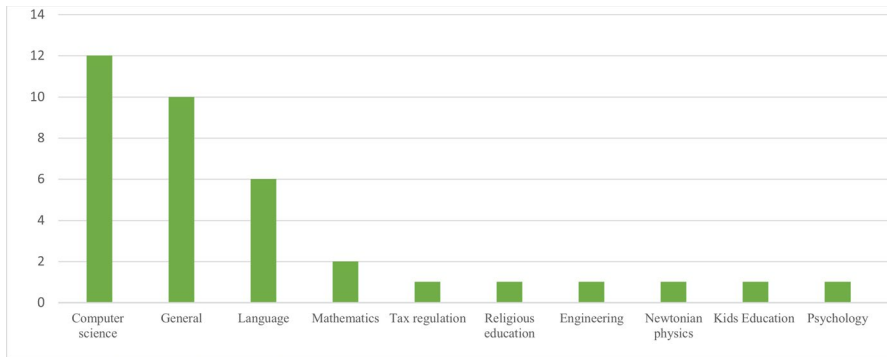


Fig. 4 The fields of the chatbots in the selected articles

topics (Fig. 4). For example, some chatbots were used as tutors for teaching programming languages such as Java (Coronado et al., 2018; Daud et al., 2020) and Python (Winkler et al., 2020), while other researchers proposed educational chatbots for computer networks (Clarizia et al., 2018; Lee et al., 2020), databases (Latham et al., 2011; Ondáš et al., 2019), and compilers (Griol et al., 2011).

Table 4. shows that ten (27.77%) articles presented general-purpose educational chatbots that were used in various educational contexts such as online courses (Song et al., 2017; Benedetto & Cremonesi, 2019; Tegos et al., 2020). The approach authors use often relies on a general knowledge base not tied to a specific field.

In comparison, chatbots used to teach languages received less attention from the community (6 articles; 16.66%). Interestingly, researchers used a variety of interactive media such as voice (Ayedoun et al., 2017; Ruan et al., 2021), video (Griol et al., 2014), and speech recognition (Ayedoun et al., 2017; Ruan et al., 2019).

A few other subjects were targeted by the educational chatbots, such as engineering (Mendez et al., 2020), religious education (Alobaidi et al., 2013), psychology (Hayashi, 2013), and mathematics (Rodrigo et al., 2012).

5.2 RQ2: What platforms do the proposed chatbots operate on?

Table 5. shows an overview of the platforms the educational chatbots operate on. Most researchers (25 articles; 69.44%) developed chatbots that operate on the web (Fig. 5). The web-based chatbots were created for various educational purposes. For example, KEMTbot (Ondáš et al., 2019) is a chatbot system that provides information about the department, its staff, and their offices. Other chatbots acted as intelligent tutoring systems, such as Oscar (Latham et al., 2011), used for teaching computer science topics. Moreover, other web-based chatbots such as EnglishBot (Ruan et al., 2021) help students learn a foreign language.

Six (16.66%) articles presented educational chatbots that exclusively operate on a mobile platform (e.g., phone, tablet). The articles were published recently in 2019 and 2020. The mobile-based chatbots were used for various purposes. Examples include Rexy (Benedetto & Cremonesi, 2019), which helps students enroll in

Table 4 An overview of the fields the chatbots targeted

Chabot Field	Number of Articles	Article(s)
Computer science	12	(Griol et al., 2011; Latham et al., 2011; Tegos et al., 2015; Benotti et al., 2017; Clarizia et al., 2018; Coronado et al., 2018; Verleger & Pembroke, 2018; Ondáš et al., 2019; Lee et al., 2020; Qin et al., 2020; Winkler et al., 2020; Daud et al., 2020)
General	10	(Redondo-Hernández & Pérez-Marín, 2011; Song et al., 2017; Schouten et al., 2017; Benedetto & Crenonesi, 2019; Janati et al., 2020; Tegos et al., 2020; Villegas-Ch et al., 2020; Wambsganss et al., 2020; Zedadra et al., 2014; da Silva Oliveira et al., 2019)
Language	6	(Griol et al., 2014; Ayedoun et al., 2017; Fryer et al., 2017; Ruan et al., 2019; Ruan et al., 2021; Wambsganss et al., 2021)
Mathematics	2	(Rodrigo et al., 2012; Matsuda et al., 2013)
Engineering	1	(Mendez et al., 2020)
Tax regulation	1	(Mellado-Silva et al., 2020)
Religious education	1	(Alobaidei et al., 2013)
Newtonian physics	1	(D'mello & Graesser, 2013)
Kids Education	1	(Law et al., 2020)
psychology	1	(Hayashi, 2013)

Table 5 Overview of the platforms the chatbots operate on

Chabot platform	Number of Articles	Article(s)
Web	25	(Griol et al., 2011; Latham et al., 2011; Griol et al., 2014; Rodrigo et al., 2012; Alobaidi et al., 2013; Zedadra et al., 2014; Tegos et al., 2015; Ayedoun et al., 2017; Fryer et al., 2017; Song et al., 2017; Clarizia et al., 2018; Coronado et al., 2018; Schouten et al., 2017; Verleger & Pembbridge, 2018; da Silva Oliveira et al., 2019; Ondáš et al., 2019; Ruan et al., 2019; Janati et al., 2020; Lee et al., 2020; Law et al., 2020; Tegos et al., 2020; Villegas-Ch et al., 2020; Winkler et al., 2020; Ruan et al., 2021; Wambsganss et al., 2021)
Mobile	6	(Benedetto & Cremonesi, 2019; Daud et al., 2020; Mellado-Silva et al., 2020; Mendez et al., 2020; Qin et al., 2020; Wambsganss et al., 2020)
Desktop	5	(D'mello & Graesser, 2013; Redondo-Hernández & Pérez-Marín, 2011; Matsuda et al., 2013; Benotti et al., 2017; Hayashi, 2013)

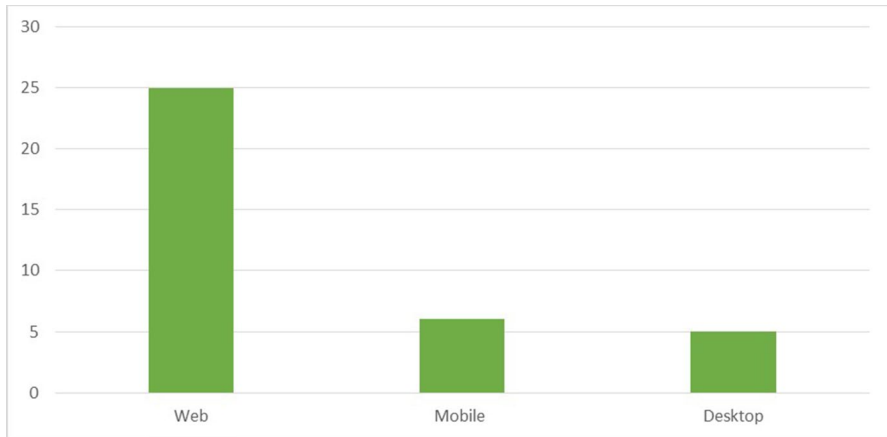


Fig. 5 The platforms of the chatbots in the selected articles

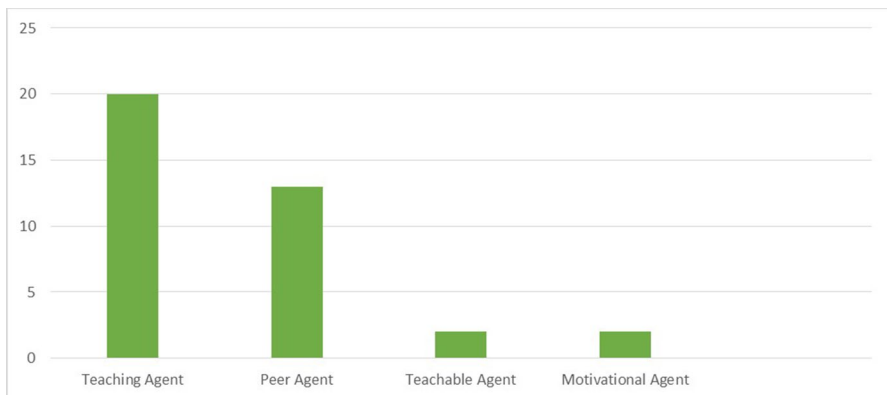


Fig. 6 The roles of the chatbots in the selected articles

courses, shows exam results, and gives feedback. Another example is the E-Java Chatbot (Daud et al., 2020), a virtual tutor that teaches the Java programming language.

Five articles (13.88%) presented desktop-based chatbots, which were utilized for various purposes. For example, one chatbot focused on the students' learning styles and personality features (Redondo-Hernández & Pérez-Marín, 2011). As another example, the SimStudent chatbot is a teachable agent that students can teach (Matsuda et al., 2013).

In general, most desktop-based chatbots were built in or before 2013, probably because desktop-based systems are cumbersome to modern users as they must be downloaded and installed, need frequent updates, and are dependent on operating systems. Unsurprisingly, most chatbots were web-based, probably because the web-based applications are operating system independent, do not require downloading,

Table 6 An overview of the roles of chatbots when interacting with students

Chabot Role	Number of Articles	Article(s)
Teaching Agent	20	(Latham et al., 2011; D'mello & Graesser, 2013; Redondo-Hernández & Pérez-Marín, 2011; Alobaidi et al., 2013; Griol et al., 2014; Zedadra et al., 2014; Ayedoun et al., 2017; Benotti et al., 2017; Song et al., 2017; Coronado et al., 2018; Ondáš et al., 2019; Ruan et al., 2019; Mellado-Silva et al., 2020; Qin et al., 2020; Villegas-Ch et al., 2020; Winkler et al., 2020; Ruan et al., 2021; Wambsganss et al., 2021; Wambsganss et al., 2020; Rodrigo et al., 2012)
Peer Agent	13	(Griol et al., 2011; Tegos et al., 2015; Fryer et al., 2017; Clarizia et al., 2018; Verleger & Pembridge, 2018; da Silva Oliveira et al., 2019; Janati et al., 2020; Lee et al., 2020; Mendez et al., 2020; Tegos et al., 2020; Hayashi, 2013; Benedetto & Cremonesi, 2019; Daud et al., 2020)
Teachable Agent	2	(Matsuda et al., 2013; Law et al., 2020)
Motivational Agent	2	(D'mello & Graesser, 2013; Schouten et al., 2017)

installing, or updating. Mobile-based chatbots are on the rise. This can be explained by users increasingly desiring mobile applications. According to an App Annie report, users spent 120 billion dollars on application stores ⁸.

5.3 RQ3 - What role do the educational chatbots play when interacting with students?

Chatbots have been found to play various roles in educational contexts, which can be divided into four roles (teaching agents, peer agents, teachable agents, and peer agents), with varying degrees of success (Table 6, Fig. 6). Exceptionally, a chatbot found in (D'mello & Graesser, 2013) is both a teaching and motivational agent.

By far, the majority (20; 55.55%) of the presented chatbots play the role of a teaching agent, while 13 studies (36.11%) discussed chatbots that are peer agents. Only two studies used chatbots as teachable agents, and two studies used them as motivational agents.

Teaching agents The teaching agents presented in the different studies used various approaches. For instance, some teaching agents recommended tutorials to students based upon learning styles (Redondo-Hernández & Pérez-Marín, 2011), students' historical learning (Coronado et al., 2018), and pattern matching (Ondáš et al., 2019). In some cases, the teaching agent started the conversation by asking

⁸ <https://www.appannie.com/en/insights/market-data/state-of-mobile-2020-infographic/>

the students to watch educational videos (Qin et al., 2020) followed by a discussion about the videos. In other cases, the teaching agent started the conversation by asking students to reflect on past learning (Song et al., 2017). Other studies discussed a scenario-based approach to teaching with teaching agents (Latham et al., 2011; D'mello & Graesser, 2013). The teaching agent simply mimics a tutor by presenting scenarios to be discussed with students. In other studies, the teaching agent emulates a teacher conducting a formative assessment by evaluating students' knowledge with multiple-choice questions (Rodrigo et al., 2012; Griol et al., 2014; Mellado-Silva et al., 2020; Wambsganss et al., 2020).

Moreover, it has been found that teaching agents use various techniques to engage students. For instance, some teaching agents engage students with a discussion in a storytelling style (Alobaidi et al., 2013; Ruan et al., 2019), whereas other chatbots engage students with effective channeling, with empathetic phrases as “uha” to show interest (Ayedoun et al., 2017). Other teaching agents provide adaptive feedback (Wambsganss et al., 2021).

Peer agents Most peer agent chatbots allowed students to ask for specific help on demand. For instance, the chatbots discussed in (Clarizia et al., 2018; Lee et al., 2020) allowed students to look up specific terms or concepts, while the peer agents in (Verleger & Pembridge, 2018; da Silva Oliveira et al., 2019; Mendez et al., 2020) were based on a Question and Answer (Q&A) database, and as such answered specific questions. Other peer agents provide more advanced assistance. For example, students may ask the peer agent in (Janati et al., 2020) how to use a particular technology (e.g., using maps in Oracle Analytics), while the peer agent described in (Tegos et al., 2015; Tegos et al., 2020; Hayashi, 2013) scaffolded a group discussion. Interestingly, the only peer agent that allowed for a free-style conversation was the one described in (Fryer et al., 2017), which could be helpful in the context of learning a language.

Teachable agents Only two articles discussed teachable agent chatbots. In general, the followed approach with these chatbots is asking the students questions to teach students certain content. For example, the chatbot discussed in (Matsuda et al., 2013) presents a mathematical equation and then asks the student of each required step to gradually solve the equation, while in the work presented in (Law et al., 2020), students individually or in a group teach a classification task to chatbots in several topics.

Motivational agents Two studies presented chatbots as motivational agent-based chatbots. One of them presented in (D'mello & Graesser, 2013) asks the students a question, then waits for the student to write an answer. Then the motivational agent reacts to the answer with varying emotions, including empathy and approval, to motivate students. Similarly, the chatbot in (Schouten et al., 2017) shows various reactionary emotions and motivates students with encouraging phrases such as “*you have already achieved a lot today*”.

Table 7 An overview of the interaction styles of the chatbots

Interaction Style	Number of Articles	Article(s)
Chatbot-driven conversation	19	(Griol et al., 2011; Latham et al., 2011; Redondo-Hernández & Pérez-Marin, 2011; Rodrigo et al., 2012; Alobaidi et al., 2013; Matsuda et al., 2013; Zedadra et al., 2014; Wambsganss et al., 2020; D'mello & Graesser, 2013; Daud et al., 2020; Griol et al., 2014; Ayedoun et al., 2017; Hayashi, 2013; Coronado et al., 2018; Schouten et al., 2017; Ruan et al., 2019; Law et al., 2020; Winkler et al., 2020; Wambsganss et al., 2021)
Intent based	13	(Benotti et al., 2017; Song et al., 2017; Clarizia et al., 2018; Verleger & Pembroke, 2018; da Silva Oliveira et al., 2019; Ondás et al., 2019; Janati et al., 2020; Lee et al., 2020; Mellado-Silva et al., 2020; Mendez et al., 2020; Qin et al., 2020; Villegas-Ch et al., 2020; Benedetto & Cremonesi, 2019)
User-driven conversation	4	(Tegos et al., 2015; Fryer et al., 2017; Tegos et al., 2020; Ruan et al., 2021)

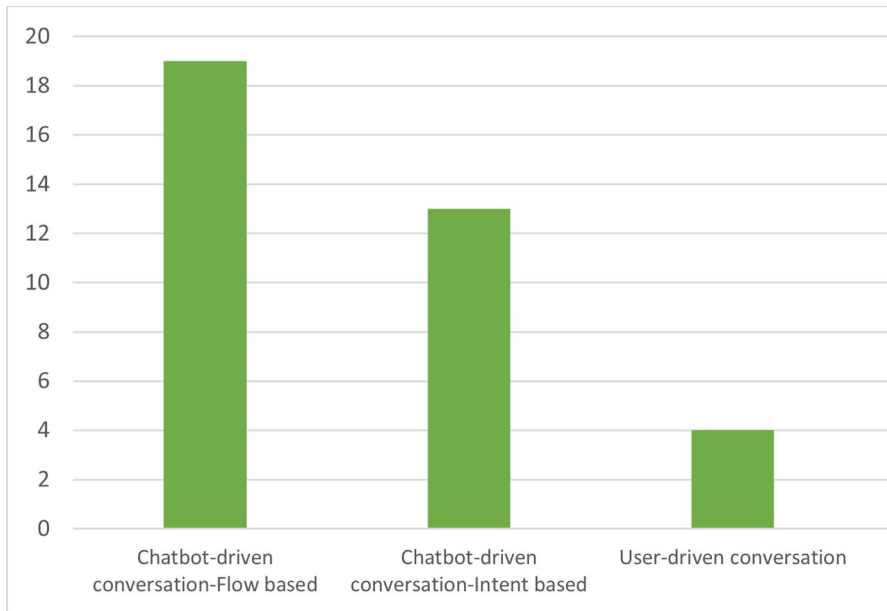


Fig. 7 The interaction styles of the chatbots in the selected articles

5.4 RQ4 – What are the interaction styles supported by the educational chatbots?

As shown in Table 7 and Fig. 7, most of the articles (88.88%) used the chatbot-driven interaction style where the chatbot controls the conversation. 52.77% of the articles used flow-based chatbots where the user had to follow a specific learning path predetermined by the chatbot. Notable examples are explained in (Rodrigo et al., 2012; Griol et al., 2014), where the authors presented a chatbot that asks students questions and provides them with options to choose from. Other authors, such as (Daud et al., 2020), used a slightly different approach where the chatbot guides the learners to select the topic they would like to learn. Subsequently, the assessment of specific topics is presented where the user is expected to fill out values, and the chatbot responds with feedback. The level of the assessment becomes more challenging as the student makes progress. A slightly different interaction is explained in (Winkler et al., 2020), where the chatbot challenges the students with a question. If they answer incorrectly, they are explained why the answer is incorrect and then get asked a scaffolding question.

The remaining articles (13 articles; 36.11%) present chatbot-driven chatbots that used an intent-based approach. The idea is the chatbot matches what the user says with a premade response. The matching could be done using pattern matching as discussed in (Benotti et al., 2017; Clarizia et al., 2018) or simply by relying on a specific conversational tool such as Dialogflow⁹ as in (Mendez et al., 2020; Lee et al., 2020; Ondáš et al., 2019).

⁹ <https://dialogflow.cloud.google.com>

Table 8 An overview of the principles used to design the chatbots

Chatbot Field	Number of Articles	Article(s)
Personalized Learning	10	(Redondo-Hernández & Pérez-Marín, 2011; Tegos et al., 2015; Griol et al., 2014; Ruan et al., 2019; Coronado et al., 2018; Janati et al., 2020; Villegas-Ch et al., 2020; Ruan et al., 2021; Wambsganss et al., 2021; Latham et al., 2011)
Experiential Learning	5	(D'mello & Graesser, 2013; Song et al., 2017; Benotti et al., 2017; Fryer et al., 2017; Qin et al., 2020)
Social Dialog	4	(Coronado et al., 2018; Qin et al., 2020; Wambsganss et al., 2021; Wambsganss et al., 2020)
Collaborative Learning	4	(Zedadra et al., 2014; Tegos et al., 2014; Tegos et al., 2020; Hayashi, 2013)
Affective Learning	2	(Ayedoun et al., 2017; Schouten et al., 2017)
Learning by teaching	2	(Matsuda et al., 2013; Law et al., 2020)
Scaffolding	1	(Winkler et al., 2020)

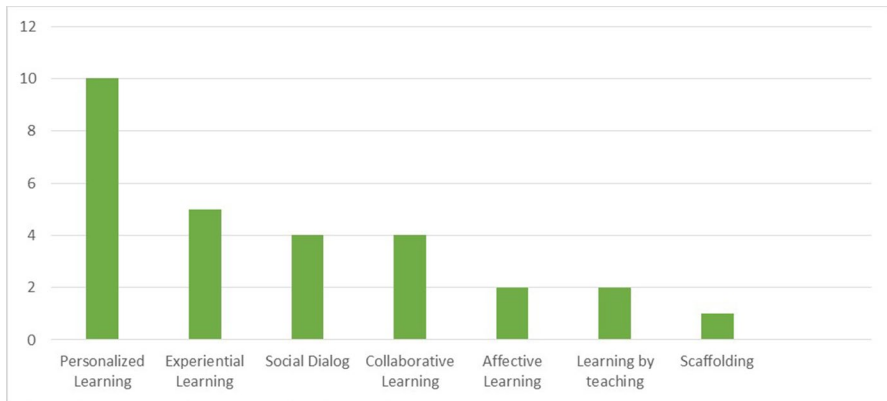


Fig. 8 The principles used to design the chatbots

Only four (11.11%) articles used chatbots that engage in user-driven conversations where the user controls the conversation and the chatbot does not have a pre-made response. For example, the authors in (Fryer et al., 2017) used Cleverbot, a chatbot designed to learn from its past conversations with humans. The authors used Cleverbot for foreign language education. User-driven chatbots fit language learning as students may benefit from an unguided conversation. The authors in (Ruan et al., 2021) used a similar approach where students freely speak a foreign language. The chatbot assesses the quality of the transcribed text and provides constructive feedback. In comparison, the authors in (Tegos et al., 2020) rely on a slightly different approach where the students chat together about a specific programming concept. The chatbot intervenes to evoke curiosity or draw students' attention to an interesting, related idea.

5.5 RQ5 – What are the principles used to guide the design of the educational chatbots?

Various design principles, including pedagogical ones, have been used in the selected studies (Table 8, Fig. 8). We discuss examples of how each of the principles was applied.

- Personalized Learning** The ability to tailor chatbots to the individual user may help meet students' needs (Clarizia et al., 2018). Many studies claim that students learn better when the chatbot is represented by a personalized method rather than a non-personalized one (Kester et al., 2005). From our selected studies, ten (27.77%) studies have applied personalized learning principles. For instance, the study in (Coronado et al., 2018) designed a chatbot to teach Java. The students' learning process is monitored by collecting information on all interactions between the students and the chatbot. Thus, direct and customized instruction and feedback are provided to students. Another notable example can be found in (Latham et al., 2011), where students were given a learning path

designed to their learning styles. With this approach, the students received 12% more accurate answers than those given chatbots without personalized learning materials. Moreover, other articles, such as the one mentioned in (Villegas-Ch et al., 2020), used AI for activity recommendation, depending on each student's needs and learning paths. The chatbot evaluates and identifies students' weaknesses and allows the AI model to be used in personalized learning.

- **Experiential Learning** Experiential learning utilizes reflection on experience and encourages individuals to gain and construct knowledge by interacting with their environment through a set of perceived experiences (Felicia, 2011). Reflection on experience is the most important educational activity for developing comprehension skills and constructing knowledge. It is primarily based on the individual's experience. Song et al. (2017) describe reflection as an intellectual activity that supports the course's weekly reflection for online learners. The chatbot asks questions to help students reflect and construct their knowledge. D'mello and Graesser (2013) have presented a constructivist view of experiential learning. The embodied chatbot mimics the conversation movements of human tutors who advise students in gradually developing explanations to problems.
- **Social Dialog** Social dialog, also called small talk, is a chit-chat that manages social situations rather than content exchange (Klüwer, 2011). The advantage of incorporating social dialog in the development of conversational agents is to establish relationships with users to engage users and gain their trust. For example, the chatbot presented in (Wambsganss et al., 2021) uses a casual chat mode allowing students to ask the chatbot to tell jokes, fun facts, or talk about unrelated content such as the weather to take a break from the main learning activity. As another example, Qin et al. (2020) suggested the usage of various social phrases that show interest, agreement, and social presence.
- **Collaborative learning** Collaborative learning is an approach that involves groups of learners working together to complete a task or solve a problem. Collaborative learning has been demonstrated to be beneficial in improving students' knowledge and improving the students' critical thinking and argumentation (Tegos et al., 2015). One of the techniques used to support collaborative learning is using an Animated Conversational Agent (ACA) (Zedadra et al., 2014). This cognitive agent considers all the pedagogical activities related to Computer-Supported Collaborative Learning (CSCL), such as learning, collaboration, and tutoring. On the other hand, the collaborative learning approach that Tegos et al. (2020) used provides an open-ended discussion, encouraging the students to work collaboratively as a pair to provide answers for a question. Before beginning the synchronous collaborative activity, the students were advised to work on a particular unit material that contained videos, quizzes, and assignments. Additionally, Tegos et al. (2015) proposed a conversational agent named MentorChat, a cloud-based CSCL, to help teachers build dialog-based collaborative activities.
- **Affective learning** Affective learning is a form of empathetic feedback given to the learner to keep the interest, attention, or desire to learn (Ayedoun et al., 2017). Two articles used this form of learning. For instance, Ayedoun et al. (2017) provided various types of affective feedback depending on the situation: congratulatory, encouraging, sympathetic, and reassuring. The idea is to sup-

port learners, mainly when a problematic situation arises, to increase their learning motivation. On the other hand, to assess the learning for low-literate people, Schouten et al. (2017) built their conversation agent to categorize four basic emotions: anger, fear, sadness, and happiness. Depending on the situation, the chatbot shows students an empathetic reaction. The researchers showed that this is helpful for learners and agents to express themselves, especially in the event of difficulty.

- **Learning by teaching** Learning by teaching is a well-known pedagogical approach that allows students to learn through the generation of explanations to others (Chase et al., 2009). Two studies used this pedagogical technique. The first study (Matsuda et al., 2013) described a chatbot that learns from students' answers and activities. Students are supposed to act as "tutors" and provide the chatbot with examples and feedback. The second study (Law et al., 2020) describes a teachable agent which starts by asking students low or high-level questions about a specific topic to evoke their curiosity. The student answers the questions, and the chatbot simulates learning. The chatbot provides a variety of questions by filling a pre-defined sentence template. To confirm its learning and make the conversation interesting, the chatbot seeks feedback from students by asking questions such as, "Am I smart?", "Am I learning?" and "Do you think I know more now than before?".
- **Scaffolding** In the educational field, scaffolding is a term describing several teaching approaches used to gradually bring students toward better comprehension and, eventually, more independence in the learning process (West et al., 2017). Teachers provide successive degrees of temporary support that aid students in reaching excellent comprehension and skill development levels that they would not be able to attain without assistance (Maybin et al., 1992). In the same way, scaffolding was used as a learning strategy in a chatbot named Sara to improve students' learning (Winkler et al., 2020). The chatbot provided voice and text-based scaffolds when needed. The approach had a significant improvement during learning in programming tasks.

5.6 RQ6 – What empirical evidence is there to substantiate the effectiveness of the proposed chatbots in education?

The surveyed articles used different types of empirical evaluation to assess the effectiveness of chatbots in educational settings. In some instances, researchers combined multiple evaluation methods, possibly to strengthen the findings.

We classified the empirical evaluation methods as follows: experiment, evaluation study, questionnaire, and focus group. An experiment is a scientific test performed under controlled conditions (Cook et al., 2002); one factor is changed at a time, while other factors are kept constant. It is the most familiar type of evaluation. It includes a hypothesis, a variable that the researcher can manipulate, and variables that can be measured, calculated, and compared. An evaluation study is a test to provide insights into specific parameters (Payne and Payne, 2004). There is typically no hypothesis to prove, and the results are often not statistically significant. A

Table 9 An overview of the applied empirical evaluation methods

Empirical evaluation	Number of Articles	Article(s)
Experiment	13	(Matsuda et al., 2013; Tegos et al., 2015; Benotti et al., 2017; Coronado et al., 2018; Schouten et al., 2017; Fryer et al., 2017; Mellado-Silva et al., 2020; Tegos et al., 2020; Winkler et al., 2020; Ruan et al., 2021; Wambsganss et al., 2021; Hayashi, 2013; Rodrigo et al., 2012)
Evaluation Study	10	(D'ello & Graesser, 2013; Redondo-Hernández & Pérez-Marín, 2011; Alobaidi et al., 2013; Ayedoun et al., 2017; Ruan et al., 2019; Janati et al., 2020; Villegas-Ch et al., 2020; Law et al., 2020; Song et al., 2017; Latham et al., 2011)
Questionnaire	10	(Griol et al., 2011; Griol et al., 2014; Zedadra et al., 2014; Clarizia et al., 2018; Benedetto & Crenonesi, 2019; Ondás et al., 2019; Daud et al., 2020; Qin et al., 2020; Wambsganss et al., 2020; da Silva Oliveira et al., 2019)
Focus group	3	(Verleger & Penbridge, 2018; Lee et al., 2020; Mendez et al., 2020)

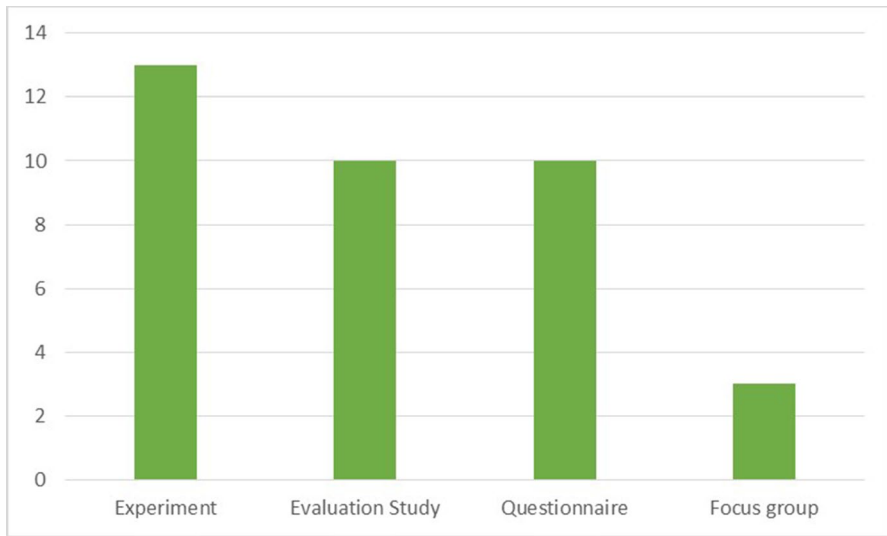


Fig. 9 Empirical evaluation methods applied in the selected studies

questionnaire is a data collection method for evaluation that focuses on a specific set of questions (Mellenbergh & Adèr, 2008). These questions aim to extract information from participants' answers. It can be carried on by mail, telephone, face-to-face interview, and online using the web or email. A focus group allows researchers to evaluate a small group or sample that represents the community (Morgan, 1996). The idea behind the focus group is to examine some characteristics or behaviors of a sample when it's difficult to examine all groups.

Table 9 and Fig. 9 show the various evaluation methods used by the articles. Most articles (13; 36.11%) used an experiment to establish the validity of the used approach, while 10 articles (27.77%) used an evaluation study to validate the usefulness and usability of their approach. The remaining articles used a questionnaire (10; 27.7%) and a focus group (3; 8.22%) as their evaluation methods.

Experiments Table 10 shows the details of the experiments the surveyed studies had used. Eight articles produced statistically significant results pointing to improved learning when using educational chatbots compared to a traditional learning setting, while a few other articles pointed to improved engagement, interest in learning, as well as subjective satisfaction.

A notable example of a conducted experiment includes the one discussed in (Wambsganss et al., 2021). The experiment evaluated whether adaptive tutoring implemented via the chatbot helps students write more convincing texts. The author designed two groups: a treatment group and a control group. The result showed that students using the chatbot (treatment group) to conduct a writing exercise wrote more convincing texts with a better formal argumentation quality than the traditional approach (control group). Another example is the experiment conducted by the authors in (Benotti et al., 2017), where the students worked on programming

Table 10 Details of experiments conducted in some reviewed articles

Article	Chatbot Approach	No. of participants	p-value	Findings
(Mellado-Silva et al., 2020)	Teaching agent, Chatbot driven	50	$p = 0.04418$	Improved learning
(Ruan et al., 2021)	Teaching agent, User driven	56	$p < 0.0001$	Improved engagement
(Fryer et al., 2017)	Peer agent, user-driven	122	$p < 0.05$	No evidence for sustained engagement
(Rodrigo et al., 2012)	Teaching agent, Chatbot driven	126	$p = 0.65$	Improved learning
(Benotti et al., 2017)	Teaching agent, chatbot driven	1000	$p = 0.01$	Improved learning, Improved interest in learning for female students
(Tegos et al., 2020)	Peer agent, User driven	56	$p = 0.03$	Improved subjective satisfaction
(Coronado et al., 2018)	Teaching agent, Chatbot driven	28	$p < 0.001$	Improved interest in learning
(Wambsganß et al., 2020)	Teaching agent, Chatbot driven	182	$p < 0.001$	Improved learning
(Wambsganß et al., 2021)	Teaching agent, Chatbot driven	55	$p < 0.05$	Improved learning
(Schouten et al., 2017)	Motivational, Chatbot driven	34	$p < 0.05$	Improved learning
(Tegos et al., 2015)	Peer agent, User driven	43	$p < 0.05$	Improved learning
(Matsuda et al., 2013)	Teachable, Chatbot driven	183	$p < 0.001$	Improved learning
(Hayashi, 2013)	Peer Agent, Chatbot driven	58	$p < 0.001$	Improved learning for text and audio-based suggestions

tasks. The experiment assessed the students' learning by a post-test. Comparing the treatment group (students who interacted with the chatbot) with a control group (students in a traditional setting), the students in the control group have improved their learning and gained more interest in learning. Another study (Hayashi, 2013) evaluated the effect of text and audio-based suggestions of a chatbot used for formative assessment. The result shows that students receiving text and audio-based suggestions have improved learning.

Despite most studies showing overwhelming evidence for improved learning and engagement, one study (Fryer et al., 2017) found that students' interest in communicating with the chatbot significantly declined in an 8-week longitudinal study where a chatbot was used to teach English.

Evaluation studies In general, the studies conducting evaluation studies involved asking participants to take a test after being involved in an activity with the chatbot. The results of the evaluation studies (Table 12) point to various findings such as increased motivation, learning, task completeness, and high subjective satisfaction and engagement.

As an example of an evaluation study, the researchers in (Ruan et al., 2019) assessed students' reactions and behavior while using 'BookBuddy,' a chatbot that helps students read books. The participants were five 6-year-old children. The researchers recorded the facial expressions of the participants using webcams. It turned out that the students were engaged more than half of the time while using BookBuddy.

Another interesting study was the one presented in (Law et al., 2020), where the authors explored how fourth and fifth-grade students interacted with a chatbot to teach it about several topics such as science and history. The students appreciated that the robot was attentive, curious, and eager to learn.

Questionnaires Studies that used questionnaires as a form of evaluation assessed subjective satisfaction, perceived usefulness, and perceived usability, apart from one study that assessed perceived learning (Table 11). Assessing students' perception of learning and usability is expected as questionnaires ultimately assess participants' subjective opinions, and thus, they don't objectively measure metrics such as students' learning.

While using questionnaires as an evaluation method, the studies identified high subjective satisfaction, usefulness, and perceived usability. The questionnaires used mostly Likert scale closed-ended questions, but a few questionnaires also used open-ended questions.

A notable example of a study using questionnaires is 'Rexy,' a configurable educational chatbot discussed in (Benedetto & Cremonesi, 2019). The authors designed a questionnaire to assess Rexy. The questionnaires elicited feedback from participants and mainly evaluated the effectiveness and usefulness of learning with Rexy. The results largely point to high perceived usefulness. However, a few participants pointed out that it was sufficient for them to learn with a human partner. One student indicated a lack of trust in a chatbot.

Table 11 Details of questionnaires conducted in some reviewed articles

Article	Chatbot Approach	No. of participants	Findings
(Griol et al., 2011)	Peer Agent, Chatbot driven	15	High subjective satisfaction
(Griol et al., 2014)	Teaching Agent, Chatbot driven	25	High ease of learning and subjective satisfaction
(Zedadra et al., 2014)	Teaching Agent, Chatbot driven	39	High subjective satisfaction
(Clarizia et al., 2018)	Peer agent, chatbot driven	167	High usability
(Benedetto & Cremonesi, 2019)	Peer agent, chatbot driven	40	High usefulness
(da Silva Oliveira et al., 2019)	Peer agent, chatbot driven	12	High subjective satisfaction
(Ondáš et al., 2019)	Teaching agent, chatbot driven	9	Moderate usefulness
(Daud et al., 2020)	Peer agent, chatbot driven	43	High usefulness
(Qin et al., 2020)	Teaching agent, chatbot driven	28	High usefulness and improved learning
(Wambsganß et al., 2020)	Teaching agent, chatbot driven	12	High usefulness

Table 12 Details of evaluation studies conducted in some reviewed articles

Article	Chatbot Approach	No. of participants	Findings
(Latham et al., 2011)	Teaching agent, chatbot driven	54	High usefulness
(D'mello & Graesser, 2013)	Motivational agent, chatbot driven	84	Improved learning
(Redondo-Hernández & Pérez-Marín, 2011)	Teaching agent, Chatbot driven	20	Moderate subjective satisfaction
(Alobaidi et al., 2013)	Teaching agent, chatbot driven	12	High usefulness
(Ayedoun et al., 2017)	Teaching agent, chatbot driven	N/A	High motivation
(Song et al., 2017)	Teaching agent, chatbot driven	11	Moderate motivation
(Ruan et al., 2019)	Teaching agent, chatbot driven	5	High engagement and usefulness
(Janati et al., 2020)	Peer agent, chatbot driven	N/A	High effectiveness
(Law et al., 2020)	Teachable agent, chatbot driven	12	Design implications
(Villegas-Ch et al., 2020)	Teaching agent, chatbot driven	N/A	Improved Task completeness

Table 13 Details of focus groups conducted in some reviewed articles

Article	Chatbot Approach	No. of participants	Findings
(Verleger & Pembridge, 2018)	Peer agent, chatbot driven	21	Lack of training
(Lee et al., 2020)	Peer agent, chatbot driven	10	High usefulness, High usability
(Mendez et al., 2020)	Peer agent, chatbot driven	10	Moderate subjective satisfaction, moderate usefulness

Another example is the study presented in (Ondáš et al., 2019), where the authors evaluated various aspects of a chatbot used in the education process, including helpfulness, whether users wanted more features in the chatbot, and subjective satisfaction. The students found the tool helpful and efficient, albeit they wanted more features such as more information about courses and departments. About 62.5% of the students said they would use the chatbot again. In comparison, 88% of the students in (Daud et al., 2020) found the tool highly useful.

Focus group Only three articles were evaluated by the focus group method. Only one study pointed to high usefulness and subjective satisfaction (Lee et al., 2020), while the others reported low to moderate subjective satisfaction (Table 13). For instance, the chatbot presented in (Lee et al., 2020) aims to increase learning effectiveness by allowing students to ask questions related to the course materials. The authors invited 10 undergraduate students to evaluate the chatbot. It turned out that most of the participants agreed that the chatbot is a valuable educational tool that

Table 14 Limitations reported in reviewed papers

Limitation	Articles
Dataset training	(Song et al., 2017; Verleger & Pembridge, 2018; Lee et al., 2020; Qin et al., 2020; Winkler et al., 2020)
User-centered design	(Verleger & Pembridge, 2018; Law et al., 2020)
Losing interest over time	(Fryer et al., 2017)
Lack of feedback	(Villegas-Ch et al., 2020)
Distractions	(Qin et al., 2020)

facilitates real-time problem solving and provides a quick recap on course material. The study mentioned in (Mendez et al., 2020) conducted two focus groups to evaluate the efficacy of chatbot used for academic advising. While students were largely satisfied with the answers given by the chatbot, they thought it lacked personalization and the human touch of real academic advisors. Finally, the chatbot discussed by (Verleger & Pembridge, 2018) was built upon a Q&A database related to a programming course. Nevertheless, because the tool did not produce answers to some questions, some students decided to abandon it and instead use standard search engines to find answers.

5.7 RQ7: What are the challenges and limitations of using proposed chatbots?

Several challenges and limitations that hinder the use of chatbots were identified in the selected studies, which are summarized in Table 14 and listed as follow:

- Insufficient or Inadequate Dataset Training** The most recurring limitation in several studies is that the chatbots are either trained with a limited dataset or, even worse, incorrectly trained. Learners using chatbots with a limited dataset experienced difficulties learning as the chatbot could not answer their questions. As a result, they became frustrated (Winkler et al., 2020) and could not wholly engage in the learning process (Verleger & Pembridge, 2018; Qin et al., 2020). Another example that caused learner frustration is reported in (Qin et al., 2020), where the chatbot gave incorrect responses.

To combat the issues arising from inadequate training datasets, authors such as (Ruan et al., 2021) trained their chatbot using standard English language examination materials (e.g., IELTS and TOEFL). The evaluation suggests an improved engagement. Further, Song et al. (2017) argue that the use of Natural Language Processing (NLP) supports a more natural conversation instead of one that relies on a limited dataset and a rule-based mechanism.

- User-centered design** User-centered design (UCD) refers to the active involvement of users in several stages of the software cycle, including requirements gathering, iterative design, and evaluation (Dwivedi et al., 2012). The ultimate goal of UCD is to ensure software usability. One of the challenges mentioned in a couple of studies is the lack of student involvement in the design process

(Verleger and Pembridge, 2018) which may have resulted in decreased engagement and motivation over time. As another example, Law et al. (2020) noted that personality traits might affect how learning with a chatbot is perceived. Thus, educators wishing to develop an educational chatbot may have to factor students' personality traits into their design.

- **Losing Interest Over Time** Interestingly, apart from one study, all of the reviewed articles report educational chatbots were used for a relatively short time. Fryer et al. (2017) found that students' interest in communicating with the chatbot significantly dropped in a longitudinal study. The decline happened between the first and the second tasks suggesting a novelty effect while interacting with the chatbot. Such a decline did not happen when students were interacting with a human partner.
- **Lack of Feedback** Feedback is a crucial element that affects learning in various environments (Hattie and Timperley, 2007). It draws learners' attention to understanding gaps and supports them gain knowledge and competencies (Narciss et al., 2014). Moreover, feedback helps learners regulate their learning (Chou & Zou, 2020). Villegas-Ch et al. (2020) noted that the lack of assessments and exercises coupled with the absence of the feedback mechanism negatively affected the chatbot's success.
- **Distractions** Usability heuristics call for a user interface that focuses on the essential elements and does not distract users from necessary information (Inostroza et al., 2012). In the context of educational chatbots, this would mean that the design must focus on the essential interactions between the chatbot and the student. Qin et al. (2020) identified that external links and popups suggested by the chatbot could be distracting to students, and thus, must be used judiciously.

6 Discussion and future research directions

The purpose of this work was to conduct a systematic review of the educational chatbots to understand their fields of applications, platforms, interaction styles, design principles, empirical evidence, and limitations.

Seven general research questions were formulated in reference to the objectives.

- **RQ1** examined the fields the educational chatbots are used in. The results show that the surveyed chatbots were used to teach several fields. More than a third of the chatbots were developed to teach computer science topics, including programming languages and networks. Fewer chatbots targeted foreign language education, while slightly less than a third of the studies used general-purpose educational chatbots. Our findings are somewhat similar to (Wollny et al., 2021), and (Hwang and Chang, 2021), although both of the review studies reported that language learning was the most targeted educational topic, followed by computer programming. Other review studies such as (Winkler & Söllner, 2018) highlighted that chatbots were used to educate students on health, well-being, and self-advocacy.

- **RQ2** identified the platforms the educational chatbots operate on. Most surveyed chatbots are executed within web-based platforms, followed by a few chatbots running on mobile and desktop platforms. The web offers a versatile platform as multiple devices can access it, and it does not require installation. Other review studies such as (Cunningham-Nelson et al., 2019) and (Pérez et al., 2020) did not discuss the platform but mentioned the tools used to develop the chatbots. Popular tools include Dialogflow¹⁰, QnA Maker¹¹, ChatFuel¹². Generally, these tools allow for chatbot deployment on web and mobile platforms. Interestingly, Winkler and Söllner (2018) highlighted that mobile platforms are popular for chatbots used for medical education.
- **RQ3** explored the roles of the chatbots when interacting with students. More than half of the surveyed chatbots were used as teaching agents that recommended educational content to students or engaged students in a discussion on relevant topics. Our results are similar to those reported in (Smutny & Schreiberova, 2020) which classified most chatbots as teaching agents that recommend content, conducted formative assessments, and set learning goals.

Slightly more than a third of the surveyed chatbots acted as peer agents which helped students ask for help when needed. Such help includes term definition, FAQ (Frequently Asked Questions), and discussion scaffolding. No studies reported the use of peer agents. However, a review study (Wollny et al., 2021) reported that some chatbots were used for scaffolding which correlates with our findings.

Two chatbots were used as motivational agents showing empathetic and encouraging feedback as students learn. A few review studies such as (Okonkwo & Ade-Ibijola, 2021) and (Winkler & Söllner, 2018) identified that chatbots are used for motivation and engagement, but no details were given.

Finally, only two surveyed chatbots acted as teachable agents where students gradually taught the chatbots.

- **RQ4** investigated the interaction styles supported by the educational chatbots. Most surveyed chatbots used a chatbot-driven conversation where the chatbot was in control of the conversation. Some of these chatbots used a predetermined path, whereas others used intents that were triggered depending on the conversation. In general, related review studies did not distinguish between intent-based or flow-based chatbots. However, a review study surveyed chatbot-driven agents that were used for FAQ (Cunningham-Nelson et al., 2019). Other review studies, such as (Winkler & Söllner, 2018) highlighted that some chatbots are flow-based. However, no sufficient details were mentioned.

Only a few surveyed chatbots allowed for a user-driven conversation where the user can initiate and lead the conversation. Other review studies reported that such chatbots rely on AI algorithms (Winkler & Söllner, 2018).

¹⁰ <https://dialogflow.cloud.google.com/>

¹¹ <https://www.qnamaker.ai/>

¹² <https://chatfuel.com/>

- **RQ5** examined the principles used to guide the design of the educational chatbots. Personalized learning is a common approach where the learning content is recommended, and instruction and feedback are tailored based on students' performance and learning styles. Most related review studies did not refer to personalized learning as a design principle, but some review studies such as (Cunningham-Nelson et al., 2019) indicated that some educational chatbots provided individualized responses to students.

Scaffolding has also been used in some chatbots where students are provided gradual guidance to help them become independent learners. Scaffolding chatbots can help when needed, for instance, when students are working on a challenging task. Other review studies such as (Wollny et al., 2021) also revealed that some chatbots scaffolded students' discussions to help their learning.

Other surveyed chatbots supported collaborative learning by advising the students to work together on tasks or by engaging a group of students in a conversation. A related review study (Winkler & Söllner, 2018) highlighted that chatbots could be used to support collaborative learning.

The remaining surveyed chatbots engaged students in various methods such as social dialog, affective learning, learning by teaching, and experiential learning. However, none of the related review studies indicated such design principles behind educational chatbots.

A few surveyed chatbots have used social dialog to engage students. For instance, some chatbots engaged students with small talk and showed interest and social presence. Other chatbots used affective learning in the form of sympathetic and reassuring feedback to support learners in problematic situations. Additionally, learning by teaching was also used by two chatbots where the chatbot acted as a student and asked the chatbot for answers and examples. Further, a surveyed chatbot used experiential learning by asking students to develop explanations to problems gradually.

- **RQ6** studied the empirical evidence used to back the validity of the chatbots. Most surveyed chatbots were evaluated with experiments that largely proved with statistical significance that chatbots could improve learning and student satisfaction. A related review study (Hwang & Chang, 2021) indicated that many studies used experiments to substantiate the validity of chatbots. However, no discussion of findings was reported.

Some of the surveyed chatbots used evaluation studies to assess the effect of chatbots on perceived usefulness and subjective satisfaction. The results are in favor of the chatbots. A related review study (Hobert & Meyer von Wolff, 2019) mentioned that qualitative studies using pre/post surveys were used. However, no discussion of findings was reported.

Questionnaires were also used by some surveyed chatbots indicating perceived subjective satisfaction, ease of learning, and usefulness. Intriguingly, a review study (Pérez et al., 2020) suggested that questionnaires were the most common method of evaluation of chatbots. Such questionnaires pointed to high user satisfaction and no failure on the chatbot's part.

Finally, only this study reported using focus groups as an evaluation method. Only three chatbots were evaluated with this method with a low number of participants, and the results showed usefulness, reasonable subjective satisfaction, and lack of training.

- **RQ7** examined the challenges and limitations of using educational chatbots. A frequently reported challenge was a lack of dataset training which caused frustration and learning difficulties. A review study (Pérez et al., 2020) hinted at a similar issue by shedding light on the complex task of collecting data to train the chatbots.

Two surveyed studies also noticed the novelty effect. Students seem to lose interest in talking to chatbots over time. A similar concern was reported by a related review study (Pérez et al., 2020).

Other limitations not highlighted by related review studies include the lack of user-centered design, the lack of feedback, and distractions. In general, the surveyed chatbots were not designed with the involvement of students in the process. Further, one surveyed chatbot did not assess the students' knowledge, which may have negatively impacted the chatbot's success. Finally, a surveyed study found that a chatbot's external links and popup messages distracted the students from the essential tasks.

The main limitation not identified in our study is chatbot ethics. A review study (Okonkwo & Ade-Ibijola, 2021) discussed that ethical issues such as privacy and trust must be considered when designing educational chatbots.

To set the ground for future research and practical implementation of chatbots, we shed some light on several areas that should be considered when designing and implementing chatbots

- **Usability Principles** Usability is a quality attribute that evaluates how easy a user interface is to use (Nathoo et al., 2019). Various usability principles can serve as guidance for designing user interfaces. For instance, Nielson presented ten heuristics considered rules of thumb¹³. Moreover, Shneiderman mentioned eight golden user interface design rules (Shneiderman et al., 2016). Further, based on the general usability principles and heuristics, some researchers devised usability heuristics for designing and evaluating chatbots (conversational user interfaces). The heuristics are based on traditional usability heuristics in conjunction with principles specific to conversation and language studies. In terms of the design phase, it is recommended to design user interfaces iteratively by involving users during the design phase (Lauesen, 2005).

The chatbots discussed in the reviewed articles aimed at helping students with the learning process. Since they interact with students, the design of the chatbots must pay attention to usability principles. However, none of the chatbots explicitly discussed the reliance on usability principles in the design phase. However, it could be argued that some of the authors designed the chatbots with usability

¹³ <https://www.nngroup.com/articles/ten-usability-heuristics/>

in mind based on some design choices. For instance, Alobaidi et al. (2013) used contrast to capture user attention, while Ayedoun et al. (2017) designed their chatbot with subjective satisfaction in mind. Further, Song et al. (2017) involved the users in their design employing participatory design, while Clarizia et al. (2018) ensured that the chatbot design is consistent with existing popular chatbots. Similarly, Villegas-Ch et al. (2020) developed the user interface of their chatbot to be similar to that of Facebook messenger.

Nevertheless, we argue that it is crucial to design educational chatbots with usability principles in mind explicitly. Further, we recommend that future educators test for the chatbot's impact on learning or student engagement and assess the usability of the chatbots.

- Chatbot Personality** Personality describes consistent and characteristic patterns of behavior, emotions, and cognition (Smestad and Volden, 2018). Research suggests that users treat chatbots as if they were humans (Chaves & Gerosa, 2021), and thus chatbots are increasingly built to have a personality. In fact, researchers have also used the Big Five model to explain the personalities a chatbot can have when interacting with users (Völkel & Kaya, 2021; McCrae & Costa, 2008). Existing studies experimented with various chatbot personalities such as agreeable, neutral, and disagreeable (Völkel & Kaya, 2021). An agreeable chatbot uses family-oriented words such as “family” or “together” (Hirsh et al., 2009), words that are regarded as positive emotionally such as “like” or “nice” (Hirsh et al., 2009), words indicating assurance such as “sure” (Nass et al., 1994), as well as certain emojis (Völkel et al., 2019), as suggested by the literature. On the other hand, a disagreeable chatbot does not show interest in the user and might be critical and uncooperative (Andrist et al., 2015).

Other personalities have also been attributed to chatbots, such as casual and formal personalities, where a formal chatbot uses a standardized language with proper grammar and punctuation, whereas a casual chatbot includes everyday, informal language (Andrist et al., 2015; Cafaro et al., 2016).

Despite the interest in chatbot personalities as a topic, most of the reviewed studies shied away from considering chatbot personality in their design. A few studies, such as (Coronado et al., 2018; Janati et al., 2020; Qin et al., 2020; Wambsganss et al., 2021), integrated social dialog into the design of the chatbot. However, the intention of the chatbots primarily focused on the learning process rather than the chatbot personality. We argue that future studies should shed light on how chatbot personality could affect learning and subjective satisfaction.

- Chatbot Localization and Acceptance** Human societies' social behavior and conventions, as well as the individuals' views, knowledge, laws, rituals, practices, and values, are all influenced by culture. It is described as the underlying values, beliefs, philosophy, and methods of interacting that contribute to a person's unique psychological and social environment. Shin et al. (2022) defines culture as the common models of behaviors and interactions, cognitive frameworks, and perceptual awareness gained via socialization in a cross-cultural environment. The acceptance of chatbots involves a cultural dimension. The cultural and social circumstances in which the chatbot is used influence how students interpret the chatbot and how they consume and engage with it. For example,

the study by (Rodrigo et al., 2012) shows evidence that the chatbot ‘Scooter’ was regarded and interacted with differently in the Philippines than in the United States. According to student gaming behavior in the Philippines, Scooter’s interface design did not properly exploit Philippine society’s demand for outwardly seamless interpersonal relationships.

Nevertheless, all other studies didn’t focus on localization as a design element crucial to the chatbot’s effectiveness and acceptance. We encourage future researchers and educators to assess how the localization of chatbots affects students’ acceptance of the chatbots and, consequently, the chatbot’s success as a learning mate.

- **Development Framework** As it currently stands, the literature offers little guidance on designing effective usable chatbots. None of the studies used a certain framework or guiding principles in designing the chatbots. Future works could contribute to the Human-Computer Interaction (HCI) and education community by formulating guiding principles that assist educators and instructional designers in developing effective usable chatbots. Such guiding principles must assist educators and researchers across multiple dimensions, including learning outcomes and usability principles. A software engineering approach can be adopted, which guides educators in four phases: requirements, design, deployment, and assessment. A conceptual framework could be devised as a result of analyzing quantitative and qualitative data from empirical evaluations of educational chatbots. The framework could guide designing a learning activity using chatbots by considering learning outcomes, interaction styles, usability guidelines, and more.
- **End-user development of chatbots** End-User Development (EUD) is a field that is concerned with tools and activities enabling end-users who are not professional software developers to write software programs (Lieberman et al., 2006). EUD uses various approaches such as visual programming (Kuhail et al., 2021) and declarative formulas (Kuhail and Lauesen, 2012). Since end-users outnumber software engineers by a factor of 30-to-1, EUD empowers a much larger pool of people to participate in software development (Kuhail et al., 2021). Only a few studies (e.g., Ondáš et al., 2019; Benedetto & Cremonesi, 2019) have discussed how the educational chatbots were developed using technologies such as Google Dialogflow and IBM Watson ¹⁴. Nevertheless, such technologies are only accessible to developers. Recently, commercial tools such as Google Dialogflow CX ¹⁵ emerged to allow non-programmers to develop chatbots with visual programming, allowing end-users to create a program by putting together graphical visual elements rather than specifying them textually.

Future studies could experiment with existing EUD tools that allow educational chatbots’ development. In particular, researchers could assess the usability and expressiveness of such tools and their suitability in the educational context.

¹⁴ <https://www.ibm.com/sa-ar/watson>

¹⁵ <https://cloud.google.com/dialogflow/cx/docs/basics>

7 Conclusion

This study described how several educational chatbot approaches empower learners across various domains. The study analyzed 36 educational chatbots proposed in the literature. To analyze the tools, the study assessed each chatbot within seven dimensions: educational field, platform, educational role, interaction style, design principles, empirical principles, and challenges as well as limitations.

The results show that the chatbots were proposed in various areas, including mainly computer science, language, general education, and a few other fields such as engineering and mathematics. Most chatbots are accessible via a web platform, and a fewer chatbots were available on mobile and desktop platforms. This choice can be explained by the flexibility the web platform offers as it potentially supports multiple devices, including laptops, mobile phones, etc.

In terms of the educational role, slightly more than half of the studies used teaching agents, while 13 studies (36.11%) used peer agents. Only two studies presented a teachable agent, and another two studies presented a motivational agent. Teaching agents gave students tutorials or asked them to watch videos with follow-up discussions. Peer agents allowed students to ask for help on demand, for instance, by looking terms up, while teachable agents initiated the conversation with a simple topic, then asked the students questions to learn. Motivational agents reacted to the students' learning with various emotions, including empathy and approval.

In terms of the interaction style, the vast majority of the chatbots used a chatbot-driven style, with about half of the chatbots using a flow-based with a predetermined specific learning path, and 36.11% of the chatbots using an intent-based approach. Only four chatbots (11.11%) used a user-driven style where the user was in control of the conversation. A user-driven interaction was mainly utilized for chatbots teaching a foreign language.

Concerning the design principles behind the chatbots, slightly less than a third of the chatbots used personalized learning, which tailored the educational content based on learning weaknesses, style, and needs. Other chatbots used experiential learning (13.88%), social dialog (11.11%), collaborative learning (11.11%), affective learning (5.55%), learning by teaching (5.55%), and scaffolding (2.77%).

Concerning the evaluation methods used to establish the validity of the approach, slightly more than a third of the chatbots used experiment with mostly significant results. The remaining chatbots were evaluated with evaluation studies (27.77%), questionnaires (27.77%), and focus groups (8.33%). The findings point to improved learning, high usefulness, and subjective satisfaction.

Some studies mentioned limitations such as inadequate or insufficient dataset training, lack of user-centered design, students losing interest in the chatbot over time, and some distractions.

There are several challenges to be addressed by future research. None of the articles explicitly relied on usability heuristics and guidelines in designing the chatbots, though some authors stressed a few usability principles such as consistency and subjective satisfaction. Further, none of the articles discussed or

assessed a distinct personality of the chatbots though research shows that chatbot personality affects users' subjective satisfaction.

Future studies should explore chatbot localization, where a chatbot is customized based on the culture and context it is used in. Moreover, researchers should explore devising frameworks for designing and developing educational chatbots to guide educators to build usable and effective chatbots. Finally, researchers should explore EUD tools that allow non-programmer educators to design and develop educational chatbots to facilitate the development of educational chatbots. Adopting EUD tools to build chatbots would accelerate the adoption of the technology in various fields.

Study Limitations We established some limitations that may affect this study. We restricted our research to the period January 2011 to April 2021. This limitation was necessary to allow us to practically begin the analysis of articles, which took several months. We potentially missed other interesting articles that could be valuable for this study at the date of submission.

We conducted our search using four digital libraries: ACM, Scopus, IEEE Xplore, and SpringerLink. We may have missed other relevant articles found in other libraries such as Web of Science.

Our initial search resulted in a total of 1208 articles. We applied exclusion criteria to find relevant articles that were possible to assess. As such, our decision might have caused a bias: for example, we could have excluded short papers presenting original ideas or papers without sufficient evidence.

Since different researchers with diverse research experience participated in this study, article classification may have been somewhat inaccurate. As such, we mitigated this risk by cross-checking the work done by each reviewer to ensure that no relevant article was erroneously excluded. We also discussed and clarified all doubts and gray areas after analyzing each selected article.

There is also a bias towards empirically evaluated articles as we only selected articles that have an empirical evaluation, such as experiments, evaluation studies, etc. Further, we only analyzed the most recent articles when many articles discussed the same concept by the same researchers.

At last, we could have missed articles that report an educational chatbot that could not be found in the selected search databases. To deal with this risk, we searched manually to identify significant work beyond the articles we found in the search databases. Nevertheless, the manual search did not result in any articles that are not already found in the searched databases.

Author Contributions Mohammad Amin Kuhail: Conceptualization, Methodology, Validation, Investigation, Data Curation, Writing, Original Draft, Writing - Review & Editing, Supervision, Project administration

Nazik Alturki: Conceptualization, Methodology, Investigation, Data Curation, Writing, Writing - Review & Editing, Supervision.

Salwa Alramlawi: Investigation, Data Curation, Writing, Formal Analysis, Writing, Writing - Review & Editing.

Kholood Alhejori: Investigation, Data Curation, Writing, Writing - Review & Editing.

Funding No funding was received to assist with the preparation of this manuscript.
The authors have no relevant financial or non-financial interests to disclose

Data Availability Statement The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- AbuShawar, B., & Atwell, E. (2015). Alice chatbot: Trials and outputs. *Computación y Sistemas*, 19(4), 625–632.
- Baylor, A.L (2011). The design of motivational agents and avatars. *Educational Technology Research and Development*, 59(2), 291–300.
- Benotti, L., Martnez, M.C., & Schapachnik, F. (2017). Atool for introducing computer science with automatic formative assessment. *IEEE Transactions on Learning Technologies*, 11(2), 179–192.
- Benotti, L., Martnez, M.C., & Schapachnik, F. (2018). Atool for introducing computer science with automatic formative assessment. *IEEE Transactions on Learning Technologies*, 11(2), 179–192. <https://doi.org/10.1109/TLT.2017.2682084>.
- Cafaro, A., Vilhjálmsón, H.H., & Bickmore, T. (2016). First impressions in human–agent virtual encounters. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 23(4), 1–40.
- Car, L.T., Dhinakaran, D.A., Kyaw, B.M., Kowatsch, T., Joty, S., Theng, Y.-L., & Atun, R. (2020). Conversational agents in health care: scoping review and conceptual analysis. *Journal of medical Internet research*, 22(8), e17158.
- Chase, C.C, Chin, D.B, Oppezzo, M.A, & Schwartz, D.L (2009). Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of Science Education and Technology*, 18(4), 334–352.
- Chaves, A.P., & Gerosa, M.A. (2021). How should my chatbot interact? a survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction*, 37(8), 729–758.
- Chou, C.-Y., & Zou, N.-B. (2020). An analysis of internal and external feedback in self-regulated learning activities mediated by self-regulated learning tools and open learner models. *International Journal of Educational Technology in Higher Education*, 17(1), 1–27.
- Cook, T.D, Campbell, D.T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. MA: Houghton Mifflin Boston.
- Coronado, M., Iglesias, C.A., Carrera, Á., & Mardomingo, A. (2018). A cognitive assistant for learning java featuring social dialogue. *International Journal of Human-Computer Studies*, 117, 55–67.
- Cunningham-Nelson, S., Baktashmotlagh, M., & Boles, W. (2019). Visualizing student opinion through-text analysis. *IEEE Transactions on Education*, 62(4), 305–311.
- Daud, S.H.M., Teo, N.H.I., & Zain, N.H.M. (2020). Ejava chatbot for learning programming language: Apost-pandemic alternative virtual tutor. *International Journal*, 8(7), 3290–3298.
- De Angeli, A., & Brahnam, S. (2008). I hate you! disinhibition with virtualpartners. *Interacting with computers*, 20(3), 302–310.
- Dehn, D.M, & Van Mulken, S. (2000). The impact of animated interface agents: a review of empirical research. *International journal of human-computer studies*, 52(1), 1–22.

- D'mello, S., & Graesser, A. (2013). Autotutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2 (4), 1–39.
- Dwivedi, M., Upadhyay, M.S., & Tripathi, A. (2012). A working framework for the user-centered design approach and a survey of the available methods. *International Journal of Scientific and Research Publications*, 2(4), 12–19.
- Fryer, L.K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners. *Computers in Human Behavior*, 75, 461–468.
- Haake, M., & Gulz, A. (2009). A look at the roles of look & roles in embodied pedagogical agents—a user preference perspective. *International Journal of Artificial Intelligence in Education*, 19(1), 39–71.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Hirsh, J.B., DeYoung, C.G., & Peterson, J.B. (2009). Meta-traits of the big five differentially predict engagement and restraint of behavior. *Journal of Personality*, 77(4), 1085–1102.
- Kester, L., Kirschner, P.A., & Van Merriënboer, J.J. (2005). The management of cognitive load during complex cognitive skill acquisition by means of computer-simulated problem solving. *British Journal of Educational Psychology*, 75(1), 71–85.
- King, F.B. (2002). A virtual student: Not an ordinary joe. *The Internet and Higher Education*, 5(2), 157–166.
- Kulik, J.A., & Fletcher, J.D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of Educational Research*, 86(1), 42–78.
- Kumar, R., & Rose, C.P. (2010). Architecture for building conversational agents that support collaborative learning. *IEEE Transactions on Learning Technologies*, 4(1), 21–34.
- Martha, A.S.D., & Santoso, H.B. (2019). The design and impact of the pedagogical agent: A systematic literature review. *Journal of Educators Online*, 16 (1), n1.
- Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Cohen, W.W., Stylianides, G.J., & Koedinger, K.R. (2013). Cognitive anatomy of tutor learning: Lessons learned with simstudent. *Journal of Educational Psychology*, 105(4), 1152.
- Morgan, D.L. (1996). Focus groups. *Annual Review of Sociology*, 22(1), 129–152.
- Narciss, S., Sosnovsky, S., Schnaubert, L., Andrès, E., Eichmann, A., Gogadze, G., & Melis, E. (2014). Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education*, 71, 56–76.
- Okonkwo, C.W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, 2, 100033.
- 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.
- Rodrigo, M.M.T., Baker, R.S., Agapito, J., Nabos, J., Repalam, M.C., Reyes, S.S., & San Pedro, M.O.C. (2012). The effects of an interactive software agent on student affective dynamics while using; an intelligent tutoring system. *IEEE Transactions on Affective Computing*, 3(2), 224–236.
- Ruttkay, Z., & Pelachaud, C. (2006). *From browstotrust: Evaluating embodied conversational agents* Vol. 7. Berlin: Springer Science & Business Media.
- Schouten, D.G., Venneker, F., Bosse, T., Neerincx, M.A., & Cremers, A.H. (2017). A digital coach that provides affective and social learning support to low-literate learners. *IEEE Transactions on Learning Technologies*, 11 (1), 67–80.
- Seel, N.M. (2011). *Encyclopedia of the sciences of learning*. Berlin: Springer Science & Business Media.
- Serenko, A., Bontis, N., & Detlor, B. (2007). End-user adoption of animated interface agent in everyday work applications. *Behaviour & Information Technology*, 26(2), 119–132.
- Shin, D., Chotiyaputta, V., & Zaid, B. (2022). The effects of cultural dimensions on algorithmic news: How do cultural value orientations affect how people perceive algorithms? *Computers in Human Behavior*, 126, 107007.
- Shneiderman, B., Plaisant, C., Cohen, M.S., Jacobs, S., Elmqvist, N., & Diakopoulos, N. (2016). *Designing the user interface: strategies for effective human-computer interaction*. London: Pearson.
- Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the facebook messenger. *Computers & Education*, 151, 103862.
- Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge (acting with technology)*. Penguin: The MIT Press.

- Tegos, S., Demetriadis, S., & Karakostas, A. (2015). Promoting academically productive talk with conversational agent interventions in collaborative learning settings. *Computers & Education*, 87, 309–325.
- Tegos, S., Demetriadis, S., & Tsiatsos, T. (2014). A configurable conversational agent to trigger students' productive dialogue: a pilot study in the call domain. *International Journal of Artificial Intelligence in Education*, 24(1), 62–91.
- VanLehn, K., Graesser, A., Jackson, G.T., Jordan, P., Olney, A., & Rosé, C.P. (2007). Natural language tutoring: A comparison of human tutors, computer tutors, and text. *Cognitive Science*, 31(1), 3–52.
- Villegas-Ch, W., Arias-Navarrete, A., & Palacios-Pacheco, X. (2020). Proposal of an architecture for the integration of a chatbot with artificial intelligence in a smart campus for the improvement of learning. *Sustainability*, 12(4), 1500.
- Walker, E., Rummel, N., & Koedinger, K.R. (2011). Designing automated adaptive support to improve student helping behaviors in a peer tutoring activity. *International Journal of Computer-Supported Collaborative Learning*, 6 (2), 279–306.
- Weizenbaum, J. (1966). Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45.
- Wik, P., & Hjalmarsson, A. (2009). Embodied conversational agents in computer assisted language learning. *Speech Communication*, 51(10), 1024–1037.
- Alkhoori, A., Kuhail, M.A., & Alkhoori, A. (2020). Unibud: A virtual academic adviser. In *2020 12th annual undergraduate research conference on applied computing (URC)* (pp. 1–4).
- Alobaidi, O.G., Crockett, K.A., O'Shea, J.D., & Jarad, T.M. (2013). Abdullah: An intelligent arabic conversational tutoring system for modern islamic education. In *Proceedings of the world congress on engineering*, Vol. 2.
- Andrist, S., Mutlu, B., & Tapus, A. (2015). Look like me: matching robot personality via gazeto increase motivation. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 3603–3612).
- Anghelescu, P., & Nicolaescu, S.V. (2018). Chatbot application using search engines and teaching methods. In *2018 10th international conference on electronics, computers and artificial intelligence (ECAI)* (pp. 1–6).
- Ayedoun, E., Hayashi, Y., & Seta, K. (2017). Communication strategies and affective back channels for conversational agents to enhance learners' willingness to communicate in a second language. In *International conference on artificial intelligence in education* (pp. 459–462).
- Benedetto, L., & Cremonesi, P. (2019). Remy, a configurable application for building virtual teaching assistants. In *IFIP conference on human-computer interaction* (pp. 233–241).
- Bradeško, L., & Mladenčić, D. (2012). A survey of chatbot systems through a loebner prize competition. In *Proceedings of Slovenian language technologies society eighth conference of language technologies* (pp. 34–37).
- Brewer, R.N., Findlater, L., Kaye, J., Lasecki, W., Munteanu, C., & Weber, A. (2018). Accessible voice interfaces. In *Companion of the 2018 ACM conference on computer supported cooperative work and social computing* (pp. 441–446).
- Budiu, R. (2018). The user experience of chatbots. *Nielsen Norman Group* 25.
- Chaudhuri, S., Kumar, R., Howley, I.K., & Rosé, C.P. (2009). Engaging collaborative learners with helping agents. In *AIED* (pp. 365–372).
- Chhibber, N., & Law, E. (2019). Using conversational agents to support learning by teaching. arXiv:1909.13443.
- Chukhno, O., Chukhno, N., Samouylov, K.E., & Shorgin, S. (2019). A chatbot as an environment for carrying out the group decision making process. In *ITMM (Selected Papers)* (pp. 15–25).
- Clarizia, F., Colace, F., Lombardi, M., Pascale, F., & Santaniello, D. (2018). Chatbot: An education support system for student. In *International symposium on cyberspace safety and security* (pp. 291–302).
- Cunningham-Nelson, S., Boles, W., Trouton, L., & Margerison, E. (2019). A review of chatbots in education: practical steps forward. In *30th annual conference for the australasian association for engineering education (AAEE 2019): educators becoming agents of change: innovate, integrate, motivate* (pp. 299–306).
- da Silva Oliveira, J., Espíndola, D.B., Barwaldt, R., Ribeiro, L.M., & Pias, M. (2019). IBM Watson application as FAQ assistant about moodle. In *2019 IEEE Frontiers in Education Conference (FIE)* (pp. 1–8).

- Dutta, D. (2017). Developing an intelligent chat-bottool to assist high school students for learning general knowledge subjects (Tech. rep.) Georgia Institute of Technology.
- Felicia, P. (2011). Handbook of research on improving learning and motivation through educational games: Multidisciplinary approaches: Multidisciplinary approaches. iGi Global.
- Feng, D., Shaw, E., Kim, J., & Hovy, E. (2006). An intelligent discussion-bot for answering student queries in threaded discussions. In *Proceedings of the 11th international conference on intelligent user interfaces* (pp. 171–177).
- Følstad, A., Skjuve, M., & Brandtzaeg, P.B. (2018). Different chatbots for different purposes: towards a typology of chatbots to understand interaction design. In *International conference on internet science* (pp. 145–156).
- Griol, D., Baena, I., Molina, J.M., & de Miguel, A.S. (2014). A multimodal conversational agent for personalized language learning. In *Ambient intelligence-software and applications* (pp. 13–21). Springer.
- Griol, D., García-Herrero, J., & Molina, J. (2011). The educagent platform: Intelligent conversational agents for e-learning applications. In *Ambient intelligence-software and applications* (pp. 117–124). Springer.
- Gulz, A., Haake, M., Silvervarg, A., Sjödn, B., & Veletsianos, G. (2011). Building a social conversational pedagogical agent: Design challenges and methodological approaches. In *Conversational agents and natural language interaction: Techniques and effective practices* (pp. 128–155). IGI Global.
- Hayashi, Y. (2013). Learner-support agents for collaborative interaction: A study on affect and communication channels.
- Heffernan, N.T., & Croteau, E.A (2004). Web-based evaluations showing differential learning for tutorial strategies employed by the ms. lindquistutor. In *International conference on intelligent tutoring systems* (pp. 491–500).
- Hobert, S., & Meyer von Wolff, R. (2019). Say hello to your new automated tutor—a structured literature review on pedagogical conversational agents.
- Hwang, G.-J., & Chang, C.-Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, 1–14.
- Inostroza, R., Rusu, C., Roncagliolo, S., Jimenez, C., & Rusu, V. (2012). Usability heuristics for touchscreen-based mobile devices. In *2012 ninth international conference on information technology-new generations* (pp. 662–667).
- Ismail, M., & Ade-Ibijola, A. (2019). Lecturer's apprentice: A chatbot for assisting novice programmers. In *2019 international multidisciplinary information technology and engineering conference (IMITEC)* (pp. 1–8).
- Janati, S.E., Maach, A., & Ghanami, D.E. (2020). Adaptive e-learning ai-powered chatbot based on multimedia indexing. *International Journal of Advanced Computer Science and Applications* 11(12). Retrieved from <https://doi.org/10.14569/IJACSA.2020.0111238>.
- Kaczorowska-Spychalska, D. (2019). Chatbots in marketing. *Management* 23(1).
- Keele, S., et al. (2007). Guidelines for performing systematic literature reviews in software engineering (Tech. Rep.) Citeseer.
- Kerlyl, A., Hall, P., & Bull, S. (2006). Bringing chatbots into education: Towards natural language negotiation of open learner models. In *International conference on innovative techniques and applications of artificial intelligence* (pp. 179–192).
- Kerry, A., Ellis, R., & Bull, S. (2008). Conversational agents in e-learning. In *International conference on innovative techniques and applications of artificial intelligence* (pp. 169–182).
- Klüwer, T. (2011). “i like your shirt”-dialogue acts for enabling socialtalk in conversational agents. In *International workshop on intelligent virtual agents* (pp. 14–27).
- Kuhail, M.A., & Lauesen, S. (2012). Customizable visualizations with formula-linked building blocks. In *GRAPP/IVAPP* (pp. 768–771).
- Kuhail, M.A., Farooq, S., Hammad, R., & Bahja, M. (2021). Characterizing visual programming approaches for end-user developers: A systematic review. *IEEE Access*.
- Latham, A., Crockett, K., McLean, D., & Edmonds, B. (2011). Oscar: an intelligent adaptive conversational agent tutoring system. In *KES international symposium on agent and multi-agent systems: technologies and applications* (pp. 563–572).
- Lauesen, S. (2005). User interface design: a software engineering perspective. Pearson Education.
- Laurillard, D. (2013). Rethinking university teaching: A conversational framework for the effective use of learning technologies. Routledge.

- Law, E., Baghaei Ravari, P., Chhibber, N., Kulic, D., Lin, S., Pantasdo, K.D., Ceha, J., Suh, S., & Dillen, N. (2020). Curiosity notebook: A platform for learning by teaching conversational agents. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–9).
- Lee, L.-K., Fung, Y.-C., Pun, Y.-W., Wong, K.-K., Yu, M.T.-Y., & Wu, N.-I. (2020). Using a multiplatform chatbot as an onlinetutor in a university course. In *2020 international symposium on educational technology (ISET)* (pp. 53–56). IEEE.
- Lieberman, H., Paternò, F., Klann, M., & Wulf, V. (2006). End-user development: An emerging paradigm. In *End user development* (pp. 1–8). Springer.
- Mabunda, K. (2020). An intelligent chatbot for guiding visitors and locating venues.
- Maybin, J., Mercer, N., & Stierer, B. (1992). Scaffolding learning in the classroom. *Thinking voices: The work of the national oracy project* 186–195.
- McCrae, R.R., & Costa, P.T.Jr. (2008). The five-factor theory of personality.
- Mellado-Silva, R., Faúndez-Ugalde, A., & Blanco-Lobos, M. (2020). Effective learning of tax regulations using different chatbot techniques.
- Mellenbergh, G.J., & Adèr, H.J. (2008). Tests and questionnaires: Construction and administration. *Advising on Research Methods: A Consultant's Companion* 211–236.
- Mendez, S., Johanson, K., Martin Conley, V., Gosha, K., A Mack, N., Haynes, C., & A Gerhardt, R. (2020). Chatbots: A tool to supplement the future faculty mentoring of doctoral engineering students. *International Journal of Doctoral Studies* 15.
- Nass, C., Steuer, J., & Tauber, E.R. (1994). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 72–78).
- Nathoo, A., Gangabissoon, T., & Bekaroo, G. (2019). Exploring the use of tangible user interfaces for teaching basic java programming concepts: A usability study. In *2019 conference on next generation computing applications (NextComp)* (pp. 1–5).
- Oh, K.-J., Lee, D., Ko, B., & Choi, H.-J. (2017). A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation. In *2017 18th IEEE international conference on mobile data management (MDM)* (pp. 371–375).
- Ondáš, S., Pleva, M., & Hládek, D. (2019). How chatbots can be involved in the education process. In *2019 17th international conference on emerging elearning technologies and applications (ICETA)* (pp. 575–580).
- Padró, L., & Stanilovsky, E. (2012). Freeing 3.0: Towards wider multilinguality. In *LREC2012*.
- Payne, G., & Payne, J. (2004). Key concepts in social research. Sage.
- Qin, C., Huang, W., & Hew, K.F. (2020). Using the community of inquiry framework to develop an educational chatbot: lesson learned from a mobile instant messaging learning environment. In *Proceedings of the 28th international conference on computers in education*.
- Ramesh, K., Ravishankaran, S., Joshi, A., & Chandrasekaran, K. (2017). A survey of design techniques for conversational agents. In *International conference on information, communication and computing technology* (pp. 336–350).
- Redondo-Hernández, A., & Pérez-Marín, D. (2011). A procedure to automatically adapt questions in student–pedagogic conversational agent dialogues. In *International conference on user modeling, adaptation, and personalization* (pp. 122–134).
- Ruan, S., Jiang, L., Xu, Q., Liu, Z., Davis, G.M., Brunskill, E., & Landay, J.A. (2021). Englishbot: An ai-powered conversational system for second language learning. In *26th international conference on intelligent user interfaces* (pp. 434–444).
- Ruan, S., Willis, A., Xu, Q., Davis, G.M., Jiang, L., Brunskill, E., & Landay, J.A. (2019). Bookbuddy: Turning digital materials into interactive foreign language lessons through a voice chatbot. In *Proceedings of the Sixth (2019) ACM conference on learning@ scale* (pp. 1–4).
- Sinha, S., Basak, S., Dey, Y., & Mondal, A. (2020). An educational chatbot for answering queries, Springer.
- Smestad, T.L., & Volden, F. (2018). Chatbot personalities matters. In *International conference on internet science* (pp. 170–181).
- Song, D., Oh, E.Y., & Rice, M. (2017). Interacting with a conversational agent system for educational purposes in online courses. In *2017 10th international conference on human system interactions (HSI)* (pp. 78–82). IEEE.
- Tegos, S., Psathas, G., Tsiatsos, T., Katsanos, C., Karakostas, A., Tsiabanis, C., & Demetriadis, S. (2020). Enriching synchronous collaboration in online courses with configurable conversational agents. In *International Conference on Intelligent Tutoring Systems* (pp. 284–294).
- Thomas, H. (2020). Critical literature review on chatbots in education.

- Verleger, M., & Pembridge, J. (2018). Apilot study integrating an ai-driven chatbot in an introductory programming course. In *2018 ieee frontiers in education conference (FIE)* (pp. 1–4).
- Völkel, S.T., Buschek, D., Pranjic, J., & Hussmann, H. (2019). Understanding emoji interpretation through user personality and message context. In *Proceedings of the 21st international conference on human-computer interaction with mobile devices and services* (pp. 1–12).
- Völkel, S.T., & Kaya, L. (2021). Examining user preference for agreeableness in chatbots. In *CUI 2021-3rd conference on conversational user interfaces* (pp. 1–6).
- Wambsganss, T., Kueng, T., Soellner, M., & Leimeister, J.M. (2021). Arguetutor: an adaptive dialog-based learning system for argumentation skills. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–13).
- Wambsganss, T., Winkler, R., Schmid, P., & Söllner, M. (2020). Designing a conversational agent as a formative course evaluation tool.
- Wambsganss, T., Winkler, R., Söllner, M., & Leimeister, J. M. (2020). A conversational agent to improve response quality in course evaluations. In *Extended Abstracts of the 2020 CHI conference on human factors in computing systems* (pp. 1–9).
- West, A., Swanson, J., & Lipscomb, L. (2017). Ch. 11 scaffolding. Instructional methods, strategies and technologies to meet the needs of all learners.
- Winkler, R., Hobert, S., Salovaara, A., Söllner, M., & Leimeister, Jan Marco (2020). Sara, the lecturer: Improving learning in online education with a scaffolding-based conversational agent. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–14).
- Winkler, R., & Söllner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. In *Academy of management annual meeting (AOM)*.
- 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.
- Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 3506–3510).
- Zedadra, A., Lafifi, Y., & Zedadra, O. (2014). Interpreting learners’traces in collaborative learning environments. In *2014 4th international symposium isko-maghreb: concepts and tools for knowledge management (isko-maghreb)* (pp. 1–8).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Mohammad Amin Kuhail¹ · Nazik Alturki² · Salwa Alramlawi³ · Kholood Alhejori⁴

Mohammad Amin Kuhail
Mohammad.Kuhail@zu.ac.ae

Salwa Alramlawi
442005464@pnu.edu.sa

Kholood Alhejori
KholoodAlhejori.cs@gmail.com

¹ College of Technological Innovation, Zayed University, Abu Dhabi 144534, United Arab Emirates

² Information System department, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia

³ College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia

⁴ Independent Scholar, Yanbu 46423, Saudi Arabia