Received: 22 March 2020 | Accepted: 13 August 2020

DOI: 10.1002/cae.22326

WILEY

REVIEW

Rediscovering the use of chatbots in education: A systematic literature review

José Quiroga Pérez[1] | Thanasis Daradoumis[1,2] | Joan Manuel Marquès Puig[1]

[1]Department of Computer Science, Multimedia and Telecommunications, Open University of Catalonia, Barcelona, Spain

[2]Department of Cultural Technology and Communication, University of Aegean University Hill, Mytilene, Greece

Correspondence
José Quiroga Pérez, Department of Computer Science, Multimedia and Telecommunications, Open University of Catalonia, Rambla Poblenou 156, 08018 Barcelona, Spain.
Email: quiroga@uoc.edu

Abstract

Chatbots have been around for years and have been used in many areas such as medicine or commerce. Our focus is on the development and current uses of chatbots in the field of education, where they can function as service assistants or as educational agents. In this research paper, we attempt to make a systematic review of the literature on educational chatbots that address various issues. From 485 sources, 80 studies on chatbots and their application in education were selected through a step-by-step procedure based on the guidelines of the PRISMA framework, using a set of predefined criteria. The results obtained demonstrate the existence of different types of educational chatbots currently in use that affect student learning or improve services in various areas. This paper also examines the type of technology used to unravel the learning outcome that can be obtained from each type of chatbots. Finally, our results identify instances where a chatbot can assist in learning under conditions similar to those of a human tutor, while exploring other possibilities and techniques for assessing the quality of chatbots. Our analysis details these findings and can provide a solid framework for research and development of chatbots for the educational field.

KEYWORDS

chatbot, conversational agent, educational bot, literature review, natural language processor

1 | INTRODUCTION

A chatbot is a tool that combines artificial intelligence (AI) and natural language processing or other technology, which enables it to interact to a certain level of conversation with a human interlocutor through text or voice. Since the creation of the first chatbot, called Eliza [73] in 1966, whose purpose is the study of natural language and communication between human beings and machine, the use of chatbots has experienced a significant growth, and there are estimates that point to increased use in the coming years [78].

Nowadays, chatbots, also known as conversational agents, conversational tutors or just bots, are occupying various areas of work in many diverse areas of science. Some chatbots perform functions related to medicine [50], assistance in telecommunications companies [12], banks [15], all thanks to the improvements as a result of the rise in AI and natural language processing techniques that have allowed a profound development of these conversational agents. Currently, chatbots are being introduced into the field of education, thanks to a process called digital transformation [46]; this transformation mainly affects some public universities, where these chatbots are used to introduce new services or improve existing ones. This approach to education is not new: studies of this process have been conducted for more than 10 years [30]. The advantages of chatbots in open

learning for schools, universities, and other learning scenarios in which repetitive practice is important are well known. In addition to answering frequently asked questions (FAQs), there are cases of chatbots whose main purpose is to serve as educational agents with the aim of relieving the workload of human teachers in their subjects or reinforcing learning in different areas. It is especially on the main advantages and disadvantages of using chatbots as teaching assistants that this research intends to focus. Being able to leave extra training or reinforcement tasks to chatbots can free human tutors from these tasks and reinforce the learning of students with time, language, or accessibility problems. As a direct consequence of this support teachers can use the time previously allotted to these tasks to carry out research or to reinforce their teaching methods.

The purpose of the study is to obtain the necessary tools to understand how chatbot technology is being implemented in educational environments, to gather information about what services chatbots offer and find out if they are tasked with improving learning in students, to see what technology is used in these environments and how it affects students, and what results are being obtained thanks to this technology. Furthermore, this study seeks to know which tools and techniques are being used in the evaluation of chatbots, as technological tools and as tools that enhance knowledge, and what quality attributes they have. It is hoped that this study will provide the necessary knowledge to design and build an education-oriented chatbot or to improve an existing one.

2 | METHODS

This research work has been carried out through a systematic review [63] of the literature. A systematic literature review is an intensive search method used to identify relevant findings on a particular research topic [53]. This study method provides information of interest to the research topic and increases the scientific value of the data obtained. The systematic literature review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework [34]. PRISMA consists of a checklist of 27 items and a four-phase flowchart, which aims to help researchers improve the reporting of systematic reviews and meta-analyses by increasing clarity and transparency, without claiming to be a quality assessment tool. Boelens et al. [9] or Van Laar et al. [69] are another examples of the use of PRISMA framework. According to this framework, an analysis must be made of the documentation recovered with each consultation and documents can only be discarded based on certain criteria. For this study, a search has been carried out for articles published in scientific journals using the following databases as reference: Scopus, Elsevier, ACM, IEEE Xplore, Web of Science, ERIC database, Wiley library, and the research group's university library. These sources have provided enough research material for this study. Moreover, other data sources such as Google Scholar and ResearchGate.net have also been taken into account. Also included are publications in journals, conference articles, and open publications, such as doctoral theses and dissertations, that can contribute meaningfully to this study. Sources from psychology and education have not been ruled out, in addition to the already obvious computer technologies, as they provide information about how learning works. At all times, a neutral position has been taken about these sources and the rest that are part of this study.

This literature review on chatbots consists of the following steps: (a) identify research questions, (b) search for relevant journal articles and conferences to resolve research issues, (c) apply PRISMA protocol to literature selection, (d) extract information from the articles and comment here, and finally (e) establish conclusions and future lines of research.

2.1 | Identify research questions

Bearing in mind that there are various types of students with diverse educational needs, and that preliminary research has yielded different types of chatbot with distinct functions which have been developed using a variety of technologies, the primary research questions are intended to obtain information about the range of chatbot types and their technology. These questions are as follows.

1. What are the different types of educational and/or educational environment chatbots currently in use?
2. How do they affect student learning or service improvement?
3. What type of technology do they use and what learning result is obtained from each of them?
4. In what cases does a chatbot help learning under conditions similar to those of a human tutor?
5. Would it be possible to evaluate the quality of chatbots, and what techniques exist for that?

The first question provides a functional approach to the current use of chatbots in education, identifying the usual roles in which they are employed. The second research question provides information about the success or failure of the use of chatbots and what problems may

be encountered. The third research question provides information about the design and technology used and whether certain technologies are directly related to the success of chatbots. The fourth question explores the current educational roles of chatbots. The fifth question addresses the concept of quality in educational chatbots and attempts to collect the mechanisms currently being used to evaluate these chatbots. Except for the first question, all of the research questions are directly related to the outcome of the learning process through the conversational tutor and will give a perspective of whether, finally, the implementation of chatbots dedicated to learning is successful and, perceived both by objective metrics and by the students themselves, is useful or not. These questions also aim to establish suitable conditions for the development of an education-oriented chatbot.

2.2 | Literature search strategy

This study has been carried out according to different information search strategies to obtain research articles and conference papers. First, to obtain relevant articles to support the first phases of the research, the databases Scopus, Science Direct, ACM, IEEE Xplore, Web of Science, ERIC database, and Wiley library were consulted from September to November 2019 and in May 2020, using the following terms: ("chatbot" or "conversational agent" or "conversational tutor" or "bot") and ("learning" or "education"). This selection of scientific databases and publications is appropriate for the search fields of the research topic of this study: computer science and education. To obtain more general references on the research topic, some articles obtained directly from the query ("chatbot" or "conversation agent" or "conversation tutor") have also been included in this study, with preference for those articles that could contribute most to the general knowledge on chatbots. These results are outside the scope of the systematic literature review. Studies published from 2005 to 2020 have been included to ensure that the technologies used by the chatbots studied have not become obsolete and that the development of the chatbots was not limited by these technologies. This range of years has been chosen to ensure that the studies have a more current view of the research issues. However, some earlier articles have been added because they serve as a reference for the subject under investigation [73], and because they offer a historical perspective to the study. Furthermore, the results have been refined by the field of study (computer science and education). It has been added as an inclusion criterion that only articles and conference papers be shown. Ultimately, for this systematic review of the literature, other preliminary articles have also been included to provide some background knowledge for the research, such as articles retrieved from Google Scholar, ResearchGate and the traditional Google search engine; final master's and doctoral thesis papers that could potentially contribute to the research have also been searched.

2.3 | Inclusion/exclusion criteria

The main inclusion criterion for research articles obtained from reliable sources (databases and scientific publications) was that they deal directly with the subject of the research in question (chatbots in education), judging by their titles, abstracts, and main contents. The following inclusion criteria were applied: (a) the study must refer to educational chatbots or chatbots used in education to improve existing services; (b) Some of the following key points about the chatbots must be discussed: the design or technology employed, their success or failure in implementation, successful improvements or problems encountered, whether they require prior training, the roles assigned to the chatbots; (c) the study must present ways to evaluate chatbots in a generic way (e.g., as technology tools) or chatbots that are educationally oriented, that is, can be evaluated as educational tools; and (d) there must be comparison between natural language processor (NPL) employed in the chatbots. The exclusion criteria were established as follows: (a) studies on chatbots used in noneducational or nonresearch related settings were excluded; (b) Studies on chatbots whose description, use or design was unclear were excluded; (c) conference papers published in a language other than English were excluded; (d) studies whose full text was not available; and (e) studies that did not have well-defined Methods, Results, or Conclusions sections. In addition, conference papers have been subjected to the same inclusion/exclusion criteria as scientific articles. The inclusion and exclusion criteria for sources external to the scientific databases took into account the presentation of the works (well-formed, with structures similar to those of the research articles compiled from the aforementioned databases) and the relevance of the authors (the number of their relevant publications). Above all, the main criterion for the inclusion of articles from external sources has been that they address the main theme of the research. There has been an increase in the number of articles related to chatbots in education in recent years. After an initial search for articles, the first selection criterion was applied, which consisted in narrowing down the search results by time range (2005–2020), selecting only the texts in English or Spanish, and discarding those articles that did not fit the main research topic or did not satisfy the defined inclusion criteria.

The second step was the elimination of duplicate articles. The third step was the screening process, where articles dealing with any of the proposed inclusion criteria were classified. After the iterative screening process, the number of articles included in this study was below 100, a handy set of articles.

3 | RESULTS

This systematic review of the literature includes 80 articles. The PRISMA process was used to identify the research questions and discussion points raised in the 80 chatbot-related articles. Figure 1 gives a summary of the iterative process of PRISMA.

3.1 | Question 1: What are the different types of educational and/or educational environment chatbots currently in use?

The systematic review of the literature has yielded some answers to this study question. Currently, chatbots used in educational settings have two different purposes.

Many chatbots used in education have a clear orientation toward service support, an area where chatbots have already found successful application in other fields (such as in medicine, banks, customer service), helping to answer FAQs. The research has yielded different types of chatbots used in the educational environment that explain the two aspects of chatbot implementation in education. Lola [37] is used to mitigate the overload of student queries during enrollment periods by the University of Murcia in Spain and Dina [54] performs the same function for the admission service of Dian Nuswantoro Semarang University, Indonesia as the Whatsapp chatbot [74] of Telkom University, Indonesia. CourseQ, CEUBot, and Differ [10] are three other examples of the potential of chatbots as assistants in the various services universities offer besides admission and registration processes. LTKABot [36] is meant for administrative tasks and promoting manual management and support for related courses, using the ChatOps paradigm (a model where people, tools, processes, and automation are connected in a transparent flow) whose mission is automation in education. FAQs Chatbot [49] is focused on answering FAQs about university communities, academic requirements from visitors, and other

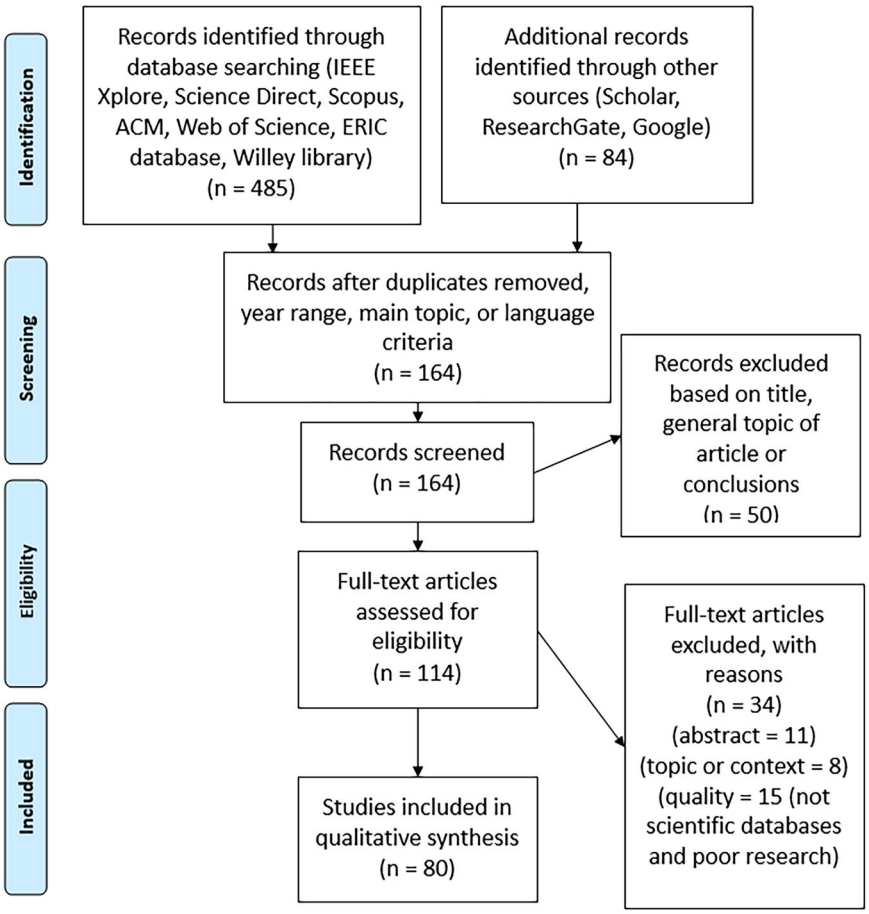

FIGURE 1 Prisma flow diagram

services. LISA [17] is a virtual assistant designed to introduce students to university life. The study tries to understand how the personality of the chatbot influences user experience and what level of intelligence should be implemented to improve the interaction. "Syllabus" Chatbot [79] answers FAQs about online postgraduate programs. LibBot [33] is a library service-oriented chatbot, which can be extrapolated to other libraries outside university environment. UCM3 Library chatbot [23] is a chatbot oriented to libraries with specific knowledge of computer science, and which combines a design adapted to mobiles with improvements in usability. FITEBot [24] is an intelligent assistant aimed at improving current services, reducing costs, and creating new services. To give better answers, it incorporates in its design an identifier of user intentions, and is configured to answer FAQs whose volume is often excessive for the administrative staff. All these chatbots may be grouped under the category of *service-oriented chatbots*.

However, there is a considerable increase in the number of chatbots whose purpose is teaching, that is, chatbots whose main purpose is to generate knowledge for specific students, usually on a specific topic. Such chatbots are used in formal education and nonformal education. The purpose of these chatbots is to generate knowledge like a human tutor would. Autotutor [22] is a multidisciplinary chatbot aimed at encouraging learning through conversation with a learner. Conversational turns are exchanged between the bot and learner. Autotutor is clearly an educational chatbot whose aim is to promote learning. Chatbot [5] is an educational chatbot aimed at promoting the study of computer science; this chatbot has been made successful through the competition "Dale aceptar." It has been remarkable in promoting computer science study among teenage girls. NerdyBot [35] is another demonstration of the power of chatbots in teaching. StudyBuddy and SmarterChild [35] are examples of chatbots used in nonformal learning and Duolingo and Ani [10] are chatbots for nonformal education. Furthermore, some chatbots are used at different educational stages, for example, ScratchThAI [29] for teaching Scratch to children, CSIEC [28] for English language learning at different educational stages, and CALMSystem [31] as a negotiated learning model for students. NDLtutor [66] has been developed to improve student involvement and reflection. Following the study of AutoTutor, Bookbuddy [52] was developed to encourage English learning by Chinese students. It uses state-of-mind detection in its operation (engagement, boredom, confusion, curiosity, happiness, frustration, or neutral). Another chatbot oriented to language learning is Mobile Chatbot [43], which provides different types of exercises to encourage learning and uses legacy aspects of gamification to maintain learner motivation [65]. Jill Watson [18] is a chatbot that attends to FAQs in the class forum of the Georgia Institute of Technology's Graduate School of Computer Science. Clive chatbot [77] is another language practice bot, whose results show that conversations are held at a good level of accuracy. Nao [61] is a humanoid robot developed with thematic maps, where topics are linked together to form a reference web. This bot, being a visible robot, helps to create a humorous and relaxed atmosphere. Interaction with students is very important for their persistence in online courses [64], where online teaching and the use of chatbots can mean that the students remain motivated until the end of the course. Student motivation is a major issue in encouraging learning. Some chatbots have specific purposes and are developed for technology certifications, such as Confucius [26], which is used as a supplement in the classroom. Xbot, a specific chatbot [2], aimed at encouraging high-school students to improve their mathematical, programming, and logic skills, highlights the motivational effect, and the success of taking them to multiple platforms (PC, mobile, tablet).

The role of educational chatbots is not limited to class assistants, they have also been tested as Virtual Patients [60], whose mission is to help train nurses to improve their effective communication skills. Schmulian and Coetzee [56] have developed chatbots visually (through the company ChatFuel) without much IT knowledge, for training in accounting. All these chatbots have been developed as educational tools and to generate knowledge. This group of chatbots can be classified as *teaching-oriented chatbots*. Within this category, we can divide them into *formal* and *nonformal* teaching-oriented chatbots. There is a summary of some chatbots found in the research in Table 1.

3.1.1 | Significant findings of the study: Ethical considerations, cultural diversity, and students with special needs

Some chatbots have been designed to educate adolescents on vital issues outside the purview of formal education. For example, Crutzen et al. [13] address complex issues during the transition of adolescents to adulthood by responding to messages about specific topics such as sex, drugs, and alcohol.

Eicher et al. [18] reveal a considerable ethical issue in chatbot technology in their study of a chatbot that answered with different degrees of accuracy to questions posed by two students; about the paternity of a male student the chatbot responded efficiently, while the answer to the same question by a female student the

TABLE 1 Subtitle: Some chatbots discovered in educational environment

Bot	Where is used	Target age of student	Service-oriented	Teaching-oriented	Reference
Lola	University of Murcia	18+	✓	–	Muñoz [37]
AutoTutor	High School	12-17	–	✓	Graesser et al. [22]
Dina	Dian N. Semarang University	18+	✓	–	Santoso et al. [54]
Chatbot	"Dale aceptar" contest	12-17	–	✓	Benotti et al. [5]
NerdyBot	WWW	–	–	✓	Molnár and Szűts [35]
StudyBuddy	WWW	–	–	✓	Molnár and Szűts [35]
SmarterChild	WWW	–	–	✓	Molnár and Szűts [35]
CourseQ	Cornell University	18+	✓	–	Brustenga et al. [10]
Differ	Norwegian Business School	18+	✓	–	Brustenga et al. [10]
LTKABot	Students of Computer Science and Electrical Engineering	18+	✓	–	Mulyana et al. [36]
CEUBot	Cardenal Herrera University	18+	✓	–	Brustenga et al. [10]
FITEBot	Science University Vietnam	18+	✓	–	Hien et al. [24]
Duolingo	WWW	–	–	✓	Brustenga et al. [10]
Whatsapp Bot	Telkom University	18+	✓	–	Wisesa and Suyanto [74]
CALMsystem	UK Students	8+	–	✓	Kerly et al. [31]
CSIEC	University, Middle School Beijing	12+	–	✓	Jia [28]
NDLTutor	University Students	18+	–	✓	Sulenam et al. [66]
ScratchThAI	Thailand Students	From children to teens	–	✓	Katchapakirin and Anutariya [29]

chatbot failed to efficiently respond to. The explanation for this may be found in the low participation of women in Science, Technology, Engineering, and Math careers, which leads to the training of AI models with subjects that mainly correspond to the mainstream, with little or no consideration for ethnic, cultural, or gender minorities. This leads to cooperation between human tutors and chatbots only as long as the chatbots are not required to attend to questions by minorities.

Cultural diversity is also an issue that affects inclusivity in education. Consejero Automático [75] is a chatbot that provides access to educational information to parents of Latino children. The cultural particularities of the participants of the study were evident. While schools pressure students to seek academic excellence, Consejero Automatico users focus on the pursuit of happiness. The study shows that the interests of this cultural group within the American society are different and that chatbots can be oriented to face different cases of cultural diversity, improving and enhancing their educational opportunities. Latinos are the largest minority in the United States, but they underperform on almost all academic fronts when compared to their African American and Caucasian counterparts. However, it has been shown that parental involvement can improve education, and this preliminary research indicated that technology may be able to address this gap, provided it is appropriately trained in social and cultural norms and values. LibBot [33] highlights the importance of socially and culturally linked dialog with students, which fosters a sense of belonging so that the students and chatbot can interact affectively. Other studies speak directly of inclusive and equitable education [38], using chatbots as a vehicle to foster an individual approach to education and encourage effective connection between teacher and student.

Accessibility is an issue that can be enhanced by chatbots [23], but these tools are not always designed with this criterion in mind which makes it difficult for people with disabilities to use them. Griol et al. [23] focused on the inclusion of students with disabilities and report benefits such as improved grades, motivation,

engagement, and development of metacognitive skills. With regard to accessibility, it is not just chatbots that can help with their integration, Bigham et al. [6] showed that the construction of a chatbot by blind students with the aim to make the tool accessible to them in addition to promoting computer science learning yielded attractive results.

3.2 | Question 2: How do chatbots affect student learning or service improvement?

Service-oriented chatbots such as Dina [54] and Lola [37] are successful in estimating the number of times they respond adequately to a service-related question. These chatbots are mainly able to evaluate this by calculating the number of satisfactory answers they give among all the questions they answer. On the other hand, teaching-oriented chatbots are meant to generate learning, so their success cannot be measured in the same way. Chatbot [5] aims to promote computer science learning, so its success or failure lies in promoting these studies to students who use this chatbot. AutoTutor [22] carries out two evaluations to measure the success of its implementation—the quality of the dialog it maintains with students and the quality of the pedagogy employed—and according to the referenced study, it achieves significant gains depending on the measure of learning sampled. ScratchThAI [29] is used to teach Scratch to students. Among other aspects of its implementation, it includes gamification to increase students' motivation. In addition, it introduces the concept of computational thinking development in students. CALMSystem [31] evaluates the results obtained by students who use the chatbot and those who do not, achieving good outcomes for this self-assessment tool. NDLtutor [66] has been evaluated by two studies that assessed two different chatbot capabilities using student questionnaires, with improvements obtained in aspects such as self-assessment and self-reflection. A study on NAO [61] concludes that although the experience is successful, interest is maintained if the conversations are interactive. FAQs Chatbot [49] produces good results when dealing with FAQs. FITEBot [24] is another chatbot that answers FAQs; it has been evaluated by means of questionnaires, with good results. LISA [17] has been evaluated through questionnaires, with positive results obtained from users. UC3M Library bot [23] has also shown good results when assessed with the aid of questionnaires, with part of the success due to its implementation on mobile devices. Another chatbot [42] for massive open online courses highlights the possible relationship between acceptance and the possibility of using mobile phones, in addition to speaking participation. Pereira et al. [42] assessed the chatbot through quantitative and qualitative analysis, obtaining good results. Other studies have sought other forms of evaluation. Galko et al. [20] defined a service-oriented chatbot for university enrollment as a terrible system for students. This study evaluated the chatbot through usability tests; it compared the time it took to register at the university by the old method and the time it took with the chatbot, obtaining a reduction in time by 26% and a reduction in errors by 77%, although the chatbot is still being tested. Jill Watson [18] has been evaluated by means of an alias without identifying the students or the human tutor to determine if the chatbot is able to serve as an assistant teacher by answering FAQs, obtaining very positive reactions and no negative reactions. With Virtual Patient [60], nursing students improved their communication skills and self-efficacy. The success of educational chatbots is predominant, and students value the positive impact [16], autonomous learning, self-organization, and self-motivation.

These evaluations of chatbots as learning tools have not used a common metric to assess the learning generated. An educational chatbot is usually assessed against the self-perceived learning of students who use the chatbot for a given subject and time interval vis-à-vis students who do not use it for the same subject and time interval. However, all of the studies referred to achieved improvements in their evaluation of the use of chatbots as educational tools.

3.2.1 | Successful, but any problems?

Although the use of service-oriented and educational chatbots has seen considerable success, certain problems that are worth mentioning have been encountered. Dibitonto et al. [17] note that the use of chatbots gives the user a false sense of anonymity and users tend to be ruder to the chatbot than they would to a human interlocutor. It emphasizes that the personality of the chatbot is essential to form an empathic relationship with the user. Shu and Riki [61] highlight the boredom that arises if there are long monologues by the chatbot and suggest that for effective communication between the chatbot and the user, the conversation should be interactive. Smutny and Schereiberova [62] have pointed out that the technology is still incipient and needs to evolve. Ciechanowski et al. [11] reported a classic problem of conversational agents, the uncanny valley, with interesting conclusions. Using psychophysiological data, they conclude that the simpler the chatbot (simpler texts and animations), the less strange and negative the conversation. Similar conclusions have been reached by Schuetzler et al. [57], who observe that the better the chatbot's

TABLE 2 Subtitle: Problems discovered

Research	Above-mentioned problem
Dibitonto et al. [17]	False sense of anonymity
Shu and Riki [61]	Boredom
Smutny and Schereiberova [62]	Incipient technology
Ciechanowski et al. [11]	Uncanny valley
Schuetzler et al. [57]	Like Uncanny valley
Molnár and Szűts [35]	Frustration
Fryer et al. [19]	Novelty effect
Eicher et al. [18]	Training side effects
Shorey et al. [60]	Technological limitations

conversational skills, the more counter-indications produced when trying to discover deceptions on the part of the interlocutor. Conversation between a student and a chatbot may lead to frustration [35] if it is not fluid enough or if the chatbot does not give the answers the student is asking for. Fryer et al. [19] found that learning with chatbots can bring about a novelty effect; this effect may lead to increased learning interest by students at the beginning of an academic year, but the interest gradually dissipates as the year progresses, which can mean loss of motivational. A chatbot, if not properly trained to account for differences and be inclusive, may give inappropriate or inadequate responses to certain issues relating to minorities. Virtual Patient [60] highlights technological limitations that make it unable to correctly identify the context of the conversation; these limitations make it to sometimes give wrong answers. There is a summary of the problems found in the research in Table 2.

3.3 | Question 3: What type of technology do chatbots use and what learning result is obtained from each of them?

Each chatbot included in this study likely differs from the rest in the design of its architecture, which is the result of the particularities that have been included in its design process. It is difficult to find two chatbots that implement exactly the same architecture, but different chatbots do have common parts that this study will simplify to give a general idea of what a "standard" chatbot architecture looks like and what parts are usually involved in its construction. The general components in the design of a chatbot are the graphical user interface (GUI), the backend and kernel, which usually takes care of NLP, and the database. This separation makes the GUI (or frontend), backend, and kernel abstract and allows for the combination of different programming languages or technologies. Villegas-Ch et al. [71] identified, in addition to the NLP core and the GUI, a component called User Experience (UX), that is responsible for making chatbot communication as natural as possible. An example of this basic scheme can be seen in Figure 2.

Chatbot [5] makes use of FreeLing (open-source software for language analysis) embedded in C++, which provides language analysis functionalities. Dina [54] employs natural language processing provided by DialogFlow (Google's NLP platform that can be used to build conversational applications in multiple languages and different platforms); besides, it contains a learning module that improves its performance with use. Smart Answering Chatbot [44] implements Artificial Intelligence Markup Language (AIML; an XML-based programming language specifically designed for the creation of the first chatbot, Eliza), a very recurrent technology in the design of chatbots [68], which Dorothy Chatbot [32] also makes

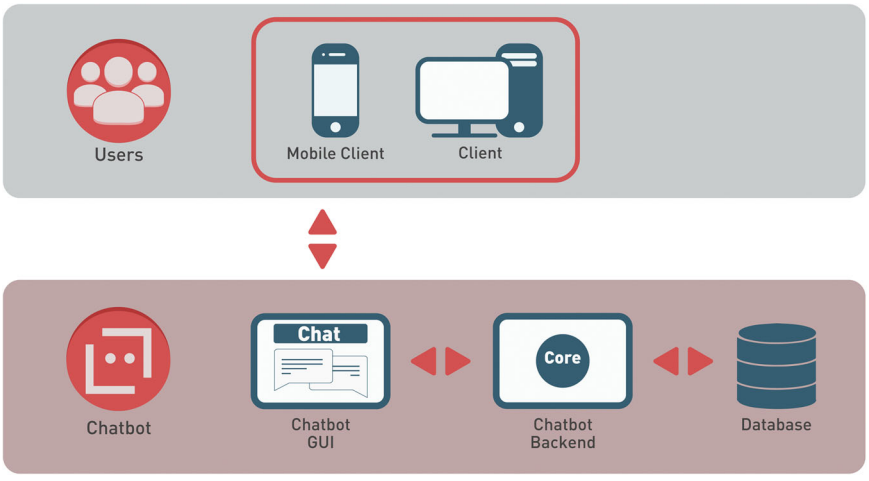

FIGURE 2 Example of chatbot architecture

use of. Lola [37] uses TensorFlow (an open-source library for machine learning developed by Google). The architecture of AutoTutor [22] is implemented with a base of C# and Visual Basic.NET, with latent semantic analysis (LSA), an NLP algorithm that relates terms and concepts, as language evaluator. Indigo [76] is developed on the platform juji.io; it is one case that deviates from the standard of the other chatbots cited and is developed in an integral way. Deep learning has been proposed as an improvement for chatbots [14], since there have been great contributions by the technology to the training of neural-based chatbot. In Table 3, we see in detail the technology used by the chatbots reviewed, as well as whether the software that uses NLP core is open source or not.

Can these technological differences affect the performance or quality of chatbots? Pirrone et al. [45] already talked about the limitations of AIML as the main technology on which some chatbots run, mentioning that these limitations can be improved upon by incorporating LSA within AIML. Shah et al. [58] support this approach. Taking as a starting point the conversational capacity, the study analyzed different noneducational chatbots (Cleverbot, Elbot, Eugene, Ultra Hal, JFred) that implement different NLP technologies and perform better than ELIZA, a chatbot whose implementation is mainly in AIML. The results of the study showed some general weaknesses in the conversational capacity of the chatbots, especially with respect to the mother language of the participants. However, many of the chatbots recognized feelings and emotions well represented in conversations. All the chatbots scored better than ELIZA in the final results, especially Cleverbot and Eugene Goostman (one of the first to pass the Turing test), which indicates that there are currently better technologies for implementing chatbots than the AIML.

3.3.1 | Training chatbots

An important aspect of designing AI-based chatbots is training. Training a chatbot to improve its conversational capabilities is a task that requires time and effort [25]. This training may involve loading sample dialogs into the chatbot database. Chatbots based on FAQs [79] are prepared with the main subjects of the course, such as required textbooks and previous forum entries, for example, FITEBot [24] has been trained with 1,560 user messages. Aujogue and Aussem [3] described the training task (7,500 handmade synthetic dialogs) as laborious and

TABLE 3 Subtitle: Examples of technology used by chatbots in education

Bot	Framework	Language processor	Open source?	Reference
Lola	Java	TensorFlow	✓	Muñoz [37]
Chatbot	C++	FreeLing	✓	Benotti et al. [5]
Dina	Node.js	DialogFlow	–	Santoso et al. [54]
Smart Answering Chatbot	HTML5 Java	AIML	✓	Pichponreay et al. [44]
AutoTutor	C# and VB.NET	LSA	–	Graesser et al. [22]
LISA	Facebook	ChatFuel	–	Dibitonto et al. [17]
FITEBot	Facebook Messenger	DialogFlow	–	Hien et al. [24]
Virtual Patient	–	DialogFlow	–	Shorey et al. [60]
FAQs Chatbot	–	AIML + LSA	–	Bhavika et al. [49]
Mobile Chatbot	Mobile devices	DialogFlow	–	Pham et al. [43]
NDLtutor	HTML5 JQuery	Php and MySQL	✓	Sulenam et al. [66]
CALMSystem	ASP.NET	LinguBot	–	Kerly et al. [31]
ScratchThAI	Open source	DialogFlow	✓	Katchapakirin and Anutariya [29]
Indigo	juji.io	juji.io	–	Xiao et al. [76]
TOB-STT	PHP	AIML	✓	Paschoal et al. [41]

after learning the chatbot brings benefits. Other chatbot implementations [40] should also be trained, proposing a set of questions and answers, where synonyms are established for identical questions or answers. However, some studies have proposed methodologies to simplify these tasks and improve the training process. Arsovski et al. [1] proposed a methodology for automatically extracting conversational knowledge and training a neural network. This study presents two stages: the first being training the chatbot from internet sources and the second being the preparation of a training dataset. The study concludes that over 75% of a chatbot's conversational knowledge can be learned, reused, and shared. Sumikawa et al. [67] collected datasets to train chatbots to answer FAQs in the context of e-learning, using as a source the questions and answers that have emerged at the Tokyo Metropolitan University during a given time interval (3 years); the study demonstrates the complexity of collecting data and training chatbots. Cuayáhuitl et al. [14] highlighted the challenge of training AI-based chatbots and the deep reinforcement learning approach as having the potential to improve chatbot training. The whole study was an attempt to improve chatbots' conversational skills by simplifying the training process and highlights the qualification of human assessors in terms of fluency, commitment, and consistency.

3.4 | Question 4: In what cases does a chatbot help learning under conditions similar to a human tutor?

A complete replacement of human teachers by chatbots cannot yet be envisaged. Some comparative studies have tried to obtain information about the knowledge transmitted by a chatbot versus by a human tutor [19], highlighting the problems of the novelty effect. A study has found that certain chatbots used for learning under very specific conditions, such as learning in foreign languages [4], can offer some advantages to the student, such as availability at any time, no mood changes, and adaptability to students with different needs. CSIEC [28] is a tool that encourages the learning of English (learning in non-mother languages is one of the main uses of chatbots in education) in an autonomous way without the assistance of any human tutor. Its evaluation has consisted in collecting diverse data such as the duration of the conversations, feedback from students and, most remarkably, the comparison of grades between students in an actual school environment that use the chatbot and those that do not, with a substantial improvement observed for those that use the chatbot as support. One of the most interesting proposals found in many of the reviewed research is the role of the conversational tutor as a teacher's assistant, where it can free teachers from having to answer certain questions that are repeated year after year, leaving them more time to focus on pure teaching while the chatbot resolves a higher percentage of the FAQs. Another area where chatbot learning works is in distance learning [39], where there is no time limitation to asking and answering questions. EduBot [70] has been put into practice in an introductory programming course, where it was used as an assistant in the development of a MATLAB internship; with the positive outcome, the authors concluded that there is great promise for chatbots in teaching. In all these cases, the chatbots have been used as assistants, with their principal function been mainly to support on certain tasks. Above all, many studies like Rooein [51] conclude that chatbots offer the opportunity for individualized education, which is where it may not be efficient to use human teachers cannot. However, Diachenko et al. [16] have noted that the collaboration of a chatbot and a human tutor is necessary, so that the human tutor can answer questions that the chatbot is unable to answer.

3.4.1 | And what do the teachers think?

This review has focused on the role of the chatbot to students. However, Bii et al. [8] note that the collaboration of teachers in the implementation of this technology is fundamental to its success. The study is focused on the Republic of Kenya, where chatbot technology is still incipient. Using questionnaires, the authors evaluated the use of chatbots among teachers, and although there were doubts raised about the use of chatbots in class, most of the teachers were positive to its implementation. Teachers stressed the importance of social awareness among students and teachers and agreed that the use of chatbots should not be used to overload the curriculum.

3.5 | Question 5: Would it be possible to evaluate the quality of chatbots, and what techniques exist for that?

The most common answer to this question is the use of questionnaires to assess the students' perception when assessing their learning with these tools. Verleger and Pembridge [70], for example, used questionnaires for this purpose. In other cases, the percentage of answers given by the chatbot that are considered correct is obtained to evaluate the effectiveness of the chatbot [54]. The importance of chatbots in education is remarkable, and

other reviews of the literature have come to that conclusion [78], highlighting, in particular, the weak ethical and educational approach being taken by the implementation of the technology. Despite this drawback, the said review highlights personalized learning as an advance, suggesting the enormous complexity of education and doubting that it can be replaced by algorithms. Wei et al. [72] focused on nontechnological attributes every chatbot should have, namely self-consciousness, humor, purity, IQ, emotional quotient, memory, self-learning, and charisma. Currently, the limitations of AI impose a clarification regarding self-consciousness, which is the identification of oneself based on details such as gender or name of the chatbot. This set of attributes form what could be defined as an artificial personality. These attributes could be a measure of the quality of a chatbot. Radziwill and Benton [48], without focusing on any special type of chatbots, argue that the attributes to measure quality should be objective and have a specific category (e.g., performance, functionality, etc.), while also suggesting the concepts of humanity and affection as quality parameters evaluable in chatbots. Other metrics used to measure the quality of a chatbot are the length and structure of the conversation [47]. If the objective of a chatbot is to teach, it seems reasonable that the measure of the quality of this chatbot be based on its teaching efficiency and not on technological criteria. Ivanovic et al. [27] introduced the term "technology-enhanced learning" for educational environments where technology supports teaching. To achieve this support, certain conditions must be met: (a) improved circumstances/environment in which educational activities can be undertaken; (b) enhanced teaching practices; (c) increased use of technology; and/or (d) improved (qualitatively/quantitatively) students' learning outcomes. In their article, Ivanovic et al. [27] discuss the incidence of chatbots as educational tools, named intelligent tutoring systems and personal agents. Even without establishing some parameters to concretely measure their quality, they set the basis to start talking about a chatbot oriented to teaching. Smutny and Schereiberova [62], after their analyses of 47 educational chatbots on Facebook Messenger, consider four main categories (teaching, humanity, affection, and accessibility) as appropriate quality attributes for teaching, even as they noted that the technology is still incipient. Savin-Baden et al. [55] define chatbots as pedagogical agents that get students to rate them as friendly, professional, or useful. According to Savin-Baden et al. [55], conversation is an important factor in building trust between a chatbot and a student, so assessing the quality of the conversation is a relevant factor in determining the quality of the chatbot. Regarding conversation, Bii and Too [7] consider the following as the attributes of a conversation: mutual understanding, common ground, trustworthiness, active listening, and humor. These are aspects to improve to enhance the conversational skills of chatbots. Mckie and Narayan [33] highlight the importance of usable, friendly, and trustworthy concepts as desirable attributes of a chatbot for students' successful use, although the study was not for a chatbot developed for teaching purposes. Hsieh [59] proposed the evaluation of chatbots using such metrics as dialog efficiency, dialog quality, and user satisfaction, although if the chatbot is service-oriented the best possible metric is whether it manages to perform the intended service or task. Gonda et al. [21] proposed the evaluation of the quality of educational chatbots based on the seven principles of good teaching: encouraging contacts between students and teacher; developing cooperation among students; encouraging active learning; prompt feedback; emphasizing time on tasks; communicating high expectations; and respect for diverse talents and ways of learning. Two main problems exist in this proposal, regardless of its interesting points, that each has to do with the absence of training for chatbots. While the architecture required to resolve these problems is complex as reported in the study, these principles are interesting and a good starting point for evaluating the quality of educational chatbots. There is a summary of the quality proposals found in the research in Table 4.

4 | DISCUSSION

This section highlights the main findings of this study: (a) Chatbots are regularly used in education in different roles, either as service assistants or as teaching assistants; finding cases of chatbots adapted to accessibility and cultural diversity, and that highlight individualized teaching is a positive aspect. (b) The success of chatbots used as service support is well proven, just like the success of teaching-oriented chatbots, regardless of their failures and problems. (c) The study shows several ways to implement a chatbot without endorsing a single formula, highlighting a basic design method and referring to the complexity of training a chatbot, with the NPL being the common feature. (d) The clearest cases of support in learning are given by chatbots whose role is to assist the teacher or to reinforce repetitive tasks. (e) There are several techniques used to evaluate the quality of a chatbot, although there is some disparity in the techniques of evaluating just the technology and the techniques of evaluating the quality of a chatbot oriented to teaching; improving a chatbot's conversation skills can help improve its quality. The limitations of this study and the implications for future research are discussed below.

TABLE 4 Subtitle: Quality attributes for chatbots and proposed approach for measure the educational quality

Research	Quality criteria	Proposed approach	Evaluating method
Verleger and Pembridge [70]	Effectiveness (perform service or task)	–	Questionnaires
Santoso et al. [54]	Do the task (perform service or task)	–	User satisfaction, Questionnaires
Zawacki-Richter et al. [78]	–	Ethical and educational approach	–
Wei et al. [72]	–	Self-consciousness, humor, EQ, IQ, and so forth	–
Radziwill and Benton [48]	Performance, functionality, humanity, affection, and so forth	–	Data analysis
Przegalinska et al. [47]	Conversation	–	
Ivanovic et al. [27]	–	TEL	–
Smutny and Schereiberova [62]	Teaching, humanity, affection, accessibility	–	Questionnaires
Savin-Baden et al. [55]	Conversation, friendly	–	Data analysis
Bii and Too [7]	Conversation, attributes of conversation	–	–
Mckie and Narayan [33]	Usable, friendly, and trustworthy	–	–
Shawar and Atwell [59]	Perform service or task	Dialog efficiency, dialog quality and user satisfaction (conversation skills)	–
Gonda et al. [21]	–	Seven principles of good teaching	–

4.1 | Summary of findings and discussion

The first research question examines chatbots that have been employed in (or are still used in) education in any capacity, including in a nonformal education setting. This study has established two well-differentiated categories of chatbots in education, namely service-oriented chatbots and teaching-oriented chatbots. Different teaching-oriented chatbots have been built and tested for different age ranges, indicating that they are versatile educational tools. In addition, teaching-oriented chatbots can assume different roles as a teacher's assistant, a complete educator, or a trainer in a certain area. The research has found a high number of teaching-oriented chatbots for language learning, which is not surprising; conversation is an effective method for learning languages, and practicing the target language at any time with a chatbot seems reasonable. The study has also found that the implementation of a chatbot on multiple platforms to enhance its use is a good idea, and that a determining factor is the motivation behind using the tool. Some chatbots has the functionality to detect the learner's mood and encourage affectivity in their response. In addition, this research has found that the use of chatbots as a pedagogical tool can help students with disabilities to advance in their studies. It can also help certain social groups and minority groups to bridge the educational "gap" that may exist between them and mainstream groups.

The second research question shows that chatbots have been successful in their areas of application. The mechanics of evaluating service-oriented chatbots is to determine, through the users' assessment of the questions answered by the chatbots, how successful they are, while evaluations of teaching-oriented chatbots are based on the learners' self-perception. This literature review suggests that the primary tool for evaluating chatbots is student questionnaires, with some exceptions such as mixed analysis or usability analysis. In other cases, the results are evaluated against tests of groups of learners who do not use the chatbot. Although the results are positive for both service-oriented and teaching-oriented chatbots and no studies were found that mention a failure of the use of a chatbot, there are some problems encountered that need to be stated, such as the novelty effect, the uncanny valley, the frustration that emerges if the participant does not get the answers he or she is

looking for, or the problems resulting from training aimed only at mainstream communities. To mitigate some of the problems identified in this study, future work on teaching-oriented chatbots should establish suitable ways to evaluate the learning provided by such chatbots. Regarding the third research question, there exists as much technology used in the development of chatbots as there are educational chatbots. Aspects such as the speed of response of a chatbot to a question and the attractiveness of its graphical user interface clearly do not depend on the programming language used in its development but on the expertise and skill of the programmer. It is clear then that the frontend and the middle layers of software used to create a chatbot do not influence the ability of the chatbot to educate. Something similar happens with the core of chatbots, usually some technology involving natural language processing is employed, but again there are many such technologies that it seems irrelevant which one to use for the development of a successful chatbot. However, there are studies that comparatively and objectively offer better results with more current technology than AIML, a technology that can be used to design a chatbot that answers FAQs but that today can be improved using other technology, for example, LSA. It should be mentioned that chatbot training is, nowadays, one of the most complex tasks in chatbot development, although there are studies that try to solve these issues by providing automated solutions that are even extendable to other chatbots in other areas.

The answer to the fourth research question shows a trend toward using teaching-oriented chatbots as support for teachers, that is, as assistants. The research has found a representative number of chatbots used as teaching assistants, with arguments that encourage their use, such as the availability to attend to questions at any time, the suitability to assist students with special abilities, the nondependence on a specific state of mind, the possibility to speak in different languages, the ability to answer FAQs, and highlighting the possibilities of improving the teaching of students with disabilities. All these add value to the use of a chatbot as a training assistant.

Concerning the fifth research question, the criteria for evaluating a chatbot can be as diverse as the works analyzed in this systematic review, mainly because of the wide array of technology that can be employed or of the NLP algorithms that can used. However, there are technological, conversational, personality, and educational criteria that provide different approaches to evaluating chatbots. All these criteria make sense and have their own importance. Evaluating a conversational agent by its conversational capabilities is reasonable, whether we are talking about technological or educational tools. The personality of a chatbot is important for establishing emotional links between a student and the chatbot, which will encourage motivation. Therefore, a chatbot with high personality skills tends to be more successful. This review has also found information about the possible evaluation of teaching-oriented chatbots with educational criteria. Further research should focus on implementing the proposals of some of the studies in that regard.

4.2 | Limitations

The first limitation observed is a disparity between the details presented by the different studies of the design and construction of their chatbots. These details may represent an inconclusive sample of the software typically used in chatbot development. Each research shows its own chatbot design and no construction protocol is applied. Another important limitation is the general lack of an environment in which chatbots have been applied for teaching purposes. Some factors can influence the success of a chatbot, such as whether it has been targeted to learners with special needs or whether the chatbot is used by online learners as against face-to-face learners. Therefore, future research should include a thorough detail of the programming languages used and a clear focus on the environment in which these learning tools are or should be applied.

4.3 | Implications

It is hoped that this review study will present both researchers and professionals with the knowledge of the uses and applications of chatbots in education. The general design plan of chatbots outlined in this review can also provide knowledge of which technologies are usually used with a reasonable degree of implementation success in chatbot development. Future researchers can also refer to this study to understand the potential that conversational tutors can have as teaching assistants at different educational levels and to obtain different evaluation criteria of these chatbots as technological tools. Notions have been provided about the problems encountered, their causes, and the effects they have on communication with chatbots. Several starting points for future research that can considerably improve these tools have also been presented, such as evaluating the ability of an education-oriented chatbot; evaluating ability with educational criteria gives better results and establishes the quality or lack thereof of these conversational tutors.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

ETHICS STATEMENT

This article does not contain any studies with human participants or animals performed by any of the authors. Human participants (university students) have merely answered an accorded questionnaire anonymously. That is, no participants' personal data were used or recorded.

ORCID

José Quiroga Pérez http://orcid.org/0000-0003-1801-303X
Thanasis Daradoumis http://orcid.org/0000-0001-9362-4469

REFERENCES

1. S. Arsovski et al., *Automatic knowledge extraction of any chatbot from conversation*, Expert Syst. With Appl. **137** (2019), 343–348. https://doi.org/10.1016/j.eswa.2019.07.014
2. W. Auccahuasi et al., *Interactive online tool as an instrument for learning mathematics through programming techniques, aimed at high school students*. Proc. 6th Int. Conf. Inf. Techn., IoT Smart City, 2018; 70–76. https://doi.org/10.1145/3301551.3301580
3. J.-B. Aujogue and A. Aussem, *Hierarchical recurrent attention networks for context-aware education chatbots*. Int. Joint Conf. Neural Netw., 2019.
4. E. Ayedoun, Y. Hayashi, and K. Seta, *A conversational agent to encourage willingness to communicate in the context of English as a foreign language*, Proc. Comput. Sci. **60** (2015), 1433–1442.
5. L. Benotti, M. C. Martínez, and F. Schapachnik, *A tool for introducing computer science with automatic formative assessment*, IEEE Trans. Learn. Technol. **11** (2018), no. 2, 179–192.
6. J. P. Bigham et al., *Inspiring blind high school students to pursue Computer Science with instant messaging chatbots*, ACM SIGCSE Bull. **40** (2008), 449–453. https://doi.org/10.1145/1352322.1352287
7. P. K. Bii and J. K. Too, *What will be in those laptops: Empowering students and teachers to add content to an educational chatbot's knowledge base*, Univer. J. Educ. Res. **4** (2016), no. 5, 941–948. https://doi.org/10.13189/ujer.2016.040502
8. P. K. Bii, J. K. Too, and C. W. Mukwa, *Teacher attitude towards use of chatbots in routine teaching*, Univer. J. Educ. Res. **6** (2018), no. 7, 1586–1597. https://doi.org/10.13189/ujer.2018.060719
9. R. Boelens, B. De Wever, and M. Voet, *Four key challenges to the design of blended learning: A systematic literature review*, Educ. Res. Rev. **22** (2017), 1–18. https://doi.org/10.1016/j.edurev.2017.06.001
10. G. G. Brustenga, M. F. Alpiste, and N. M. Castells, *Briefing paper: Chatbots in education*. 2018, available at http://hdl.handle.net/10609/80185
11. L. Ciechanowski et al., *In the shades of the uncanny valley: An experimental study of human–chatbot interaction*, Future Gener. Comput. Syst. **92** (2019), 539–548. https://doi.org/10.1016/j.future.2018.01.055
12. *Conoce a Aura, el asistente conversacional de Telefónica*. 2019, available at https://planetachatbot.com/aura-asistente-conversacional-telefonica-e0f8bd04731
13. R. Crutzen et al., *An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: An exploratory study*, J. Adolesc. Health **48** (2011), 514–519. https://doi.org/10.1016/j.jadohealth.2010.09.002
14. H. Cuayáhuitl et al., *Ensemble-based deep reinforcement learning for chatbots*, Neurocomputing **366** (2019), 118–130.
15. *Descubre cuáles son los chatbots de los bancos que te ayudarán con tus finanzas*. 2019, available at https://www.elespanol.com/invertia/mis-finanzas/20190225/descubre-chatbots-bancos-ayudaran-finanzas/378963604_0.html
16. A. V. Diachenko et al., *The use of innovative pedagogical technologies for automation of the specialists' professional training*, Int. J. Higher Educ. **8** (2019), no. 6, 288. https://doi.org/10.5430/ijhe.v8n6p288
17. M. Dibitonto et al., *Chatbot in a campus environment: Design of LiSA, a virtual assistant to help students in their university life*, Human-computer interaction. Interaction technologies Lecture Notes in Computer Science **10903**, Springer, Cham, Switzerland.
18. B. Eicher, L. Polepeddi, and A. Goel, *Jill Watson doesn't care if you're pregnant: Grounding AI ethics in empirical studies*. Proc. AAAI/ACM Conf. AI, Ethics, Soc., 2018: 88–94. https://doi.org/10.1145/3278721.3278760
19. L. K. Fryer et al., *Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners*, Comput. Human. Behav **75** (2017), 461–468.
20. L. Galko, J. Porubän, and J. Senko, *Improving the user experience of electronic university enrolment*. Int. Conf. Emerg. e-Learn. Technol. Appl., 2018. https://doi.org/10.1109/ICETA.2018.8572054
21. D. E. Gonda et al., *Evaluation of developing educational chatbots based on the Seven Principles for Good Teaching*. IEEE Int. Conf. Teach., Assessment, Learn. Eng., 2018. https://doi.org/10.1109/TALE.2018.8615175
22. A. C. Graesser et al., *AutoTutor: An intelligent tutoring system with mixed-initiative dialogue*, IEEE Trans. Educ. **48** (2005), no. 4, 612–618.
23. D. Griol, J. M. Molina, and Z. Callejas, *Incorporating android conversational agents in m-learning apps*, Expert Syst. **34** (2017), no. 4, e12156. https://doi.org/10.1111/exsy.12156
24. H. T. Hien et al., *Intelligent assistants in higher-education environments: The FITEBot, a chatbot for administrative and learning support*. 9th Int. Symp. Inf. Commun. Technol., 2018. https://doi.org/10.1145/3287921.3287937
25. G. Hiremath et al., *Chatbot for education system*, Int. J. Adv. Res., Ideas Innov. Technol. **4** (2018), no. 3, 13.
26. S.-W. Hsieh, *Effects of cognitive styles on an MSN virtual learning companion system as an adjunct to classroom instruction*, Educ. Technol. Soc. **14** (2011), no. 2, 161–174.
27. M. Ivanovića et al., *Experiences and perspectives of technology-enhanced learning and teaching in higher education—Serbian case*, Proc. Comput. Sci. **126** (2018), 1351–1359. https://doi.org/10.1016/j.procs.2018.08.086

28. J. Jia, *CSIEC: A computer assisted English learning chatbot based on textual knowledge and reasoning*, Knowl.-Based Syst. **22** (2009), 249–255. https://doi.org/10.1016/j.knosys.2008.09.001
29. K. Katchapakirin and C. Anutariya, *An architectural design of ScratchThAI: A conversational agent for Computational Thinking Development using Scratch*. 10th Int. Conf. Adv. Inf. Technol., 2018. https://doi.org/10.1145/3291280.3291787
30. A. Kerly, P. Hall, and S. Bull, *Bringing chatbots into education: Towards natural language negotiation of open learner models*, Knowl.-Based Syst. **20** (2007), 177–185. https://doi.org/10.1016/j.knosys.2006.11.014
31. A. Kerly, R. Ellis, and S. Bull, *CALMsystem: A conversational agent for learner modelling*, Knowl.-Based Syst. **21** (2008), 238–246. https://doi.org/10.1016/j.knosys.2007.11.015
32. M. D. Leonhardt et al., *Using chatbots for network management training through problem-based oriented education*. 7th IEEE Int. Conf. Adv. Learn. Technol., 2007. https://doi.org/10.1109/ICALT.2007.275
33. I. A. S. Mckie and B. Narayan, *Enhancing the academic library experience with chatbots: An exploration of research and implications for practice*, J. Australian Lib. Inf. Assoc. **68** (2019), no. 3, 268–277. https://doi.org/10.1080/24750158.2019.1611694
34. D. Moher et al., *Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement*, PLoS Med. **6** (2009), 1000097. https://doi.org/10.1371/journal.pmed.1000097.
35. G. Molnár and Z. Szűts, *The role of chatbots in formal education*. 16th Int. Symp. Intell. Syst. Inform., 2018. https://doi.org/10.1109/SISY.2018.8524609
36. E. Mulyana, R. Hakimiy, and T. Hendrawan, *Bringing automation to the classroom: A ChatOps-based approach*. 4th Int. Conf. Wireless Telematics, 2018. https://doi.org/10.1109/ICWT.2018.8527810
37. A. P. Muñoz, *Lola, el chatbot inteligente que triunfa entre los estudiantes*. 2018, available at https://retina.elpais.com/retina/2018/11/30/innovacion/1543580663_865121.html.
38. N. Nenkov et al., *Intelligent and cognitive technologies in education of International Economic relations students and Human Resource Development: Methodology in language teaching and distance learning*, Eur. J. Sustain. Develop. **6** (2017), no. 4, 353–360. https://doi.org/10.14207/ejsd.2017.v6n4p353
39. A. J. M. Neto and M. A. Fernandes, *Chatbot and conversational analysis to promote collaborative learning in distance education*. IEEE 19th Int. Conf. Adv. Learn. Technol., 2019. https://doi.org/10.1109/ICALT.2019.00102
40. K. Palasundram et al., *Sequence to sequence model performance for education chatbot*, Int. J. Emerg. Technol. Learn. **14** (2019), no. 24, 56. https://doi.org/10.3991/ijet.v14i24.12187
41. L. N. Paschoal et al., Towards a conversational agent to support the software testing education. Proc. XXXIII Brazilian Symp. Softw. Eng., 2019. https://doi.org/10.1145/3350768.3352456
42. J. Pereira et al., *Promoting learners' voice productions using chatbots as a tool for improving the learning process in a MOOC*, Technol., Knowl. Learn. **24** (2019), 545–565. https://doi.org/10.1007/s10758-019-09414-9
43. X. L. Pham et al., *Chatbot as an intelligent personal assistant for mobile language learning*. Proc. 2nd Int. Conf. Educ. E-Learn. Assoc. Comput. Machine. 2018. https://doi.org/10.1145/3291078.3291115
44. L. Pichponreay et al., *Smart answering chatbot based on OCR and overgenerating transformations and ranking*. 8th Int. Conf. Ubiquitous Future Netw., 2016. https://doi.org/10.1109/ICUFN.2016.7536948
45. R. Pirrone et al., *Semantics driven interaction using natural language in students tutoring*, Knowledge-based intelligent information and engineering systems Lecture Notes in Computer Science **4694**, Springer, Berlin, Garmany, 2007, pp. 720–727. https://doi.org/10.1007/978-3-540-74829-8_88
46. *Plan de transformación digital*. 2019, available at https://digital.um.es/plan-de-transformacion-digital/
47. A. Przegalinska et al., *In bot we trust: A new methodology of chatbot performance measures*, Bus. Horiz. **62** (2019), 785–797. https://doi.org/10.1016/j.bushor.2019.08.005
48. N. Radziwill and M. Benton, *Evaluating quality of chatbots and intelligent conversational agents*. 2017, available at https://www.researchgate.net/publication/316184347_Evaluating_Quality_of_Chatbots_and_Intelligent_Conversational_Agents.
49. B. R. Ranoliya, N. Raghuwanshi, and S. Singh, *Chatbot for university Related FAQs*. Int. Conf. Adv. Comput., Commun. Inform., 2017. https://doi.org/10.1109/ICACCI.2017.8126057
50. S. Roca et al., *Microservice chatbot architecture for chronic patient support*, J. Biomed. Inf. **102** (2019), 103305. https://doi.org/10.1016/j.jbi.2019.103305
51. D. Rooein, *Data-Driven edu chatbots*. Compan. Proc. World Wide Web Conf., 2019: 46–49. https://doi.org/10.1145/3308560.3314191
52. S. Ruan et al., *BookBuddy: Turning digital materials into interactive foreign language lessons through a voice chatbot*. ACM Conf. Learn., 2019. https://doi.org/10.1145/3330430.3333643
53. T. M. Ruiz et al., *Requirements and constructors for tailoring software processes: A systematic literature review*, Softw. Qual. J. **20** (2012), 229–260. https://doi.org/10.1007/s11219-011-9147-6
54. H. A. Santoso et al., *Dinus Intelligent Assistance (DINA) chatbot for university admission services*. Int. Seminar Appl. Technol. Inf. Commun., 2018: 417–423. https://doi.org/10.1109/ISEMANTIC.2018.8549797
55. M. Savin-Baden, G. Tombs, and R. Bhakta, *Beyond robotic wastelands of time: Abandoned pedagogical agents and new pedalled pedagogies*, E-Learn. Digit. Media **12** (2015), no. 3-4, 295–314. https://doi.org/10.1177/2042753015571835
56. A. Schmulian and S. A. Coetzee, *The development of Messenger bots for teaching and learning and accounting students' experience of the use thereof*, Br. J. Educ. Technol. **50** (2019), no. 5, 2751–2777. https://doi.org/10.1111/bjet.12723
57. R. M. Schuetzler, G. M. Grimes, and J. S. Giboney, *The effect of conversational agent skill on user behavior during deception*, Comput. Human Behav., **97** 2019, pp. 250–259. https://doi.org/10.1016/j.chb.2019.03.033
58. H. Shah et al., *Can machines talk? Comparison of Eliza with modern dialogue Systems*, Comput. Human Behav. **58** (2016), 278–295. https://doi.org/10.1016/j.chb.2016.01.004
59. B. A. Shawar and E. S. Atwell, *Different measurements metrics to evaluate a chatbot system.* Proc. Workshop Bridging Gap: Acad. Ind. Res Dialog Technol., 2007: 89–96. https://dl.acm.org/doi/10.5555/1556328.1556341
60. S. Shorey et al., *A virtual counseling application using artificial intelligence for communication skills training in nursing*

education: Development study, J. Med. Int. Res. **21** (2019), no. 10, 14658. https://doi.org/10.2196/14658

61. M. Shu, and I. Riki, *Chatbot and dialogue demonstration with a humanoid robot in the lecture class*, Universal access in human–computer interaction. Human and technological environments Lecture Notes in Computer Science **10279**, Springer, Cham, Switzerland.
62. P. Smutny, and P. Schreiberova, *Chatbots for learning: A review of educational chatbots for the Facebook Messenger*, Comput. Educ. **151** (2020), 103862. https://doi.org/10.1016/j.compedu.2020.103862
63. H. Snyder, *Literature review as a research methodology: An overview and guidelines*, J. Bus. Res. **104** (2019), 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039
64. D. Song, E. Y. Oh, and M. Rice, *Interacting with a conversational agent system for educational purposes in online courses*. 10th Int. Conf. Human Syst. Interact., 2017. https://doi.org/10.1109/HSI.2017.8005002
65. K. Stoeffler et al., *Gamified performance assessment of collaborative problem solving skills*, Comput. Human. Behav. **104** (2020), 106036. https://doi.org/10.1016/j.chb.2019.05.033
66. R. M. Suleman, R. Mizoguchi, and M. A. Ikeda, *New perspective of negotiation-based dialog to enhance metacognitive skills in the context of open learner models*, Int. J. Artif. Intell. Educ. **26** (2016), 1069–1115. https://doi.org/10.1007/s40593-016-0118-8
67. Y. Sumikawa et al., *An FAQ dataset for E-learning system used on a Japanese university*, Data Brief **25** (2019), 104001. https://doi.org/10.1016/j.dib.2019.104001
68. R. Sutoyo et al., *Designing an emotionally realistic chatbot framework to enhance its believability with AIML and information states*, Proc. Comput. Sci. **157** (2019), 621–628.
69. E. Van Laar et al., *The relation between 21st-century skills and digital skills: A systematic literature review*, Comput. Human. Behav. **72** (2017), 577–588. https://doi.org/10.1016/j.chb.2017.03.010
70. M. Verleger and J. Pembridge, *A pilot study integrating an AI-driven chatbot in an introductory programming course*. IEEE Frontiers Educ. Conf., 2018. https://doi.org/10.1109/FIE.2018.8659282
71. W. Villegas-Ch, A. Arias-Navarrete, and X. Palacios-Pacheco, *Proposal of an architecture for the integration of a chatbot with artificial intelligence in a smart campus for the improvement of learning*, Sustainability **12** (2020), 1500. https://doi.org/10.3390/su12041500
72. C. Wei, Z. Yu, and S. Fong, *How to build a chatbot: Chatbot framework and its capabilities*. Proc. 10th Int. Conf. Mach. Learn. Comput., 2018: 369–373. https://doi.org/10.1145/3195106.3195169
73. J. Weizenbaum, *ELIZA—A computer program for the study of natural language communication between man and machine*, Commun. ACM **9** (1966). https://doi.org/10.1145/365153.365168
74. C. Y. Wisesa and S. Suyanto, *Indonesian chatbot of university admission using a question answering system based on sequence-to-sequence model*, Proc. Comput. Sci. **157** (2019), 367–374. https://doi.org/10.1016/j.procs.2019.08.179
75. M. Wong-Villacres et al., *Consejero automatico: Chatbots for supporting Latino parents' educational engagement*. Proc. 10th Int. Conf. Inf. Commun. Technol. Develop., 2019: 1–5. https://doi.org/10.1145/3287098.3287149
76. Z. Xiao, M. X. Zhou, and W.-T. Fu, *Who should be my teammates: Using a conversational agent to understand individuals and help teaming*. 24th Int. Conf. Intell. User Interfaces, 2019. https://doi.org/10.1145/3301275.3302264
77. J. Zakos, and L. Capper, *CLIVE—An artificially intelligent chat robot for conversational language practice*, Artificial intelligence: Theories, models and applications Lecture Notes in Computer Science **5138**, Springer, Berlin, Heidelberg, 2008.
78. O. Zawacki-Richter, V. I. Marín, and M. Bond, *Systematic review of research on artificial intelligence applications in higher education—Where are the educators?* Int. J. Educ. Technol Higher Educ. **16** (2019), 39. https://doi.org/10.1186/s41239-019-0171-0
79. V. Zoroayka, *Design and implementation of a chatbot in online higher education settings*, Issues Inf. Syst. **19** (2018), no. 4, 44–52.

AUTHOR BIOGRAPHIES

José Quiroga Pérez holds a Master's degree in Computer Science from the Open University of Catalonia. He combines his role as a PhD candidate at the Open University of Catalonia with his work as an analyst-programmer at the IT services of the University of Murcia. His research interests include distributed systems, artificial intelligence, natural language processing, and education.

Thanasis Daradoumis holds a PhD degree in Computer Science from the Polytechnic University of Catalonia. He combines his role as an Associate Professor at the Department of Cultural Technology and Communication, University of the Aegean, with his collaboration at the Open University of Catalonia within the Department of Computer Science, Multimedia and Telecommunications. He is also a researcher at the Internet Interdisciplinary Institute (IN3) as well as at the eLearn Center (eLC). His research interests are: emotional intelligence, alternative (holistic) education, learning analytics, e-learning, collaborative, affective and adaptive systems, CSCL. He has been an advisor of over 20 MSc theses and 7 PhD already completed. He has published over 110 Scopus-indexed papers, from which over 20 are articles in ISI-JCR journals. He is a member of the editorial board of several international conferences and journals, whereas he has coordinated or participated in various national and European R&D projects.

Joan Manuel Marquès Puig holds a PhD degree in Computer Science from the Polytechnic University of Catalonia. He is an Associate Professor at Computer Sciences, Multimedia & Telecommunication Studies at the Open University of Catalonia (UOC) since 1997. His research interests include the design of scalable and cooperative Internet services and applications, privacy, and e-learning.

How to cite this article: Quiroga Pérez J, Daradoumis T, Marquès Puig JM. Rediscovering the use of chatbots in education: A systematic literature review. *Comput Appl Eng Educ*. 2020;1–17. https://doi.org/10.1002/cae.22326