International Journal of Educational Research 122 (2023) 102233

Contents lists available at ScienceDirect

International Journal of Educational Research

journal homepage: www.elsevier.com/locate/ijedures

Here's to the future: Conversational agents in higher education- a scoping review

Daniela S.M. Pereira [a,b,c], Filipe Falcão [a,b,c], Lilian Costa [a,b,c], Brian S. Lunn [d], José Miguel Pêgo [a,b,c], Patrício Costa [a,b,*]

[a] Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Largo do Paço, Braga 4700-000, Portugal
[b] ICVS/3B's, PT Government Associate Laboratory, Largo do Paço, 4700-000, Braga, Portugal
[c] iCognitus4ALL – IT Solutions, Braga 4710-057, Portugal
[d] School of Medical Education, The Medical School, Newcastle University, Newcastle upon Tyne, England

ARTICLE INFO

Keywords:
Chatbots
Higher education
Natural language processing
Technology-enhanced learning

ABSTRACT

With AI's advancing technology, pedagogical changes are occurring, and chatbots are becoming more intertwined in our daily lives. While these can be used in a variety of disciplines, they play a particularly significant role in the digital transformation of education. We present a scoping review of chatbots in higher education (HE), investigating where chatbots are being applied. The main benefits and challenges of chatbots were explored. The quality of the selected papers was assessed using the International Organization for Standards (ISO) 9241–11 guidelines. A total of 2692 citations were found while searching recognized digital databases. After reading the abstract and full text, 64 publications were considered. Backward and forward reference checking yielded four further studies. Altogether, 66 studies were included. This review provides a comprehensive overview of research on using chatbots in HE, including advantages, and challenges. Research demonstrates the versatility and the promising aspects of this type of support system for university education.

1. Introduction

1.1. Rationale

We increasingly live in an era of Artificial Intelligence (AI), where people can interact with AI, often without being aware, anytime, anywhere. Society and AI coexist, and although AI is the focal point of research and innovation, it has become commonplace and an everyday subject matter (Adamopoulou and Moussiades, 2020; Sandu & Gide, 2019).

Understanding the distinction between weak AI and strong AI is necessary to comprehend how AI is being implemented in our daily life (Flowers, 2018; Sandu & Gide, 2019). Weak AI uses AI techniques like data mining and machine learning (ML), which refers to attempts to model the human mind in the same manner that, for example, the weather phenomena are modelled. By extension, it does not strive to recreate or produce a mind in the same way that a computer model of a storm does not seek to reproduce an actual storm (Flowers, 2018; Sandu & Gide, 2019). On the other hand, strong AI refers to the development of adaptable machines that can solve

* Corresponding author at: Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho Campus de Gualtar, 4710-057, Braga, Portugal.
E-mail address: pcosta@med.uminho.pt (P. Costa).

https://doi.org/10.1016/j.ijer.2023.102233
Received 15 February 2023; Received in revised form 31 July 2023; Accepted 14 August 2023
Available online 7 September 2023

problems in the same way that the human mind can, implying that it is more than just a tool for studying the mind; rather, a properly programmed computer is a mind (Flowers, 2018; Liu, 2021).

Although strong AI is still in the initial stages, it is expected to advance in the following decades. Most current AI breakthroughs focus on weak AI; however, integrating AI into educational settings requires the usage of strong AI (Sandu & Gide, 2019). Until recently, no close form of strong AI existed, until the arrival of LLM generative chatbots such as ChatGPT (Generative Pre-trained Transformer) (Zhai, 2023). This innovative AI model attracted a lot of interest for its versatility in a range of natural language tasks (Kung et al., 2023).

AI is the key driver of change in almost every industry, including education. AI is one of the most significant technological trends, and it is expected to impact education in the future, according to the 2020 EDUCAUSE Horizon report (Brown et al., 2020). With AI's advancing technology, pedagogical changes are occurring constantly, and among all the AI applications, chatbots are becoming more and more intertwined into our daily lives (Dolianiti et al., 2020). While these can be used in a variety of disciplines, they play a particularly significant role in the digital transformation of education, providing students and educators with tools (Brown et al., 2020; Dolianiti et al., 2020; Paek & Kim, 2021; Wang, Saha, Gregori, Joyner, & Goel, 2021; Zain, 2021). Changes in the educational technology landscape require a more personalized experience in both online and face-to-face interactions to support individual learning patterns and attend to the diverse demands of each student (Dolianiti et al., 2020). Educators are now faced with the problem of keeping up with this generation's needs. Chatbots may be the answer since they provide an opportunity to communicate with younger generations in a native way (Gonda & Chu, 2019; Ondáš, Pleva, & Hládek, 2019).

1.2. Chatbots - Fundamentals

In 1950, Alan Turing generated the idea of chatbots by speculating if a group of people could talk to a computer program without realizing that it was simulated (Turing, 1950), the origin of the so-called Turing test (Adamopoulou and Moussiades, 2020).

Chatbots are AI-based software that can simulate a conversation with a human user using natural language (Clarizia, Colace, Lombardi, Pascale, & Santaniello, 2018; Dolianiti et al., 2020; Sandu & Gide, 2019) by integrating multiple data sources (such as databases or knowledge bases- KB). They are considered an exemplar utilization of ML in education and are becoming a popular tool for students (Ceha et al., 2021; Dolianiti et al., 2020; Meyer von Wolff, Nörtemann, Hobert, & Schumann, 2020).

Generally, chatbots can be classified regarding different parameters, such as knowledge domain, service provided, goals and response generation method (Adamopoulou and Moussiades, 2020; Sandu & Gide, 2019).

The knowledge domain classification is dependent on the information the chatbot accesses or the amount of data it is trained on. There are two types of domains, open and closed. Closed domain chatbots focus on a certain knowledge subject and may fail to answer other questions, but open domain chatbots can successfully discuss numerous topics (Adamopoulou and Moussiades, 2020; Sandu & Gide, 2019). Closed domain chatbots are easier to design and usually produce good outcomes, whereas open domain chatbots are still difficult to build and generate many false positive results (Lokman & Ameedeen, 2019).

Based on their services, chatbots can be divided into three categories: interpersonal, intrapersonal, and inter-agent (Sandu & Gide, 2019). The chatbots are categorized based on the sentimental proximity to the user, the degree of personal connection that occurs, and the task it is asked to carry out (Nimavat & Champaneria, 2017). Interpersonal chatbots offer services such as restaurant and flight reservations. They are facilitators since they get information and pass it to the user, which can be friendly, but they are not expected to be so (Adamopoulou and Moussiades, 2020). Intrapersonal chatbots have a major role in the users' personal life since they manage calendars and appointments (Tamrakar & Wani, 2021). These might not be dominant in the current scenario, but as natural language understanding (NLU) progresses, they will become more prominent (Nimavat & Champaneria, 2017). Lastly, inter-agent chatbots require that two systems communicate to accomplish a task. The Alexa-Cortana integration is one example of an inter-agent design (Nimavat & Champaneria, 2017; Tamrakar & Wani, 2021).

Chatbots can also be classified according to their end goal (Adamopoulou and Moussiades, 2020). Goal-based chatbots can be

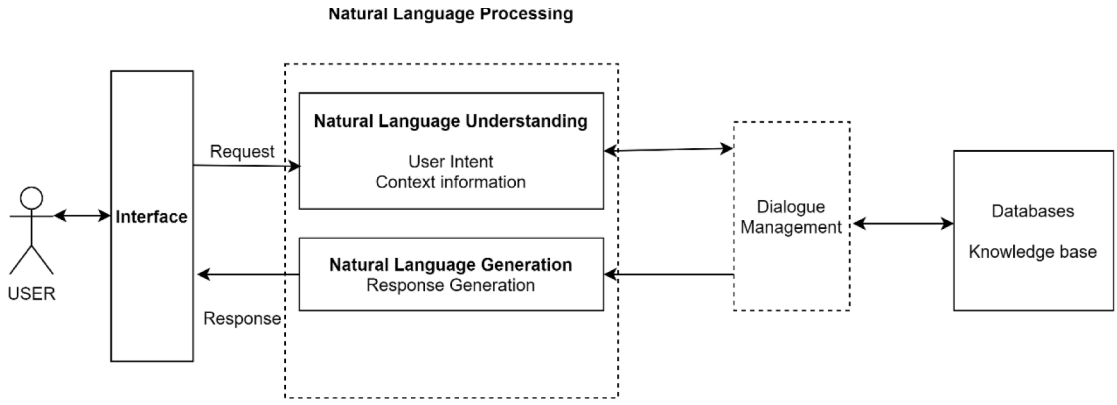

Fig. 1. The general chatbot Structure. (Adapted from: Campillos-Llanos, Thomas, Bilinski, Zweigenbaum, & Rosset, 2019; Meyer et al., 2020).

further classified into informative, conversational and task-based chatbots (Sandu & Gide, 2019). As the name suggests, Informative chatbots are designed to deliver information to the user (for example, FAQ chatbots) (Nimavat & Champaneria, 2017). Conversational chatbots attempt to communicate with the user as if they were a real person and aim to respond appropriately to the user's demands. As a result, they adopt cross-questioning, evasion, and politeness strategies to promote the user's interaction (Tamrakar & Wani, 2021). Task-based chatbots perform a specific task, such as booking a flight or helping somebody (Adamopoulou and Moussiades, 2020).

Regarding the input method and the responses generated, chatbots can be classified as rule-based and hybrid systems. While some chatbots allow input in natural language and process and generate output in that language, others, known as rule-based, process input according to rules (Sandu & Gide, 2019). Hybrid systems integrate rules such as algorithms alongside ML. An example of this case is when an outline flowchart regulates discussion direction, while natural language processing (NLP) is used to respond (Tamrakar & Wani, 2021).

However, it is crucial to note that chatbots do not have to fit into only one of these categories. Classification is helpful in informing the user about what to expect from a bot, narrowing the gap between user expectations and the chatbot output (Nimavat & Champaneria, 2017). Fig. 1 describes the simplified architecture of a chatbot.

The design of a chatbot is usually composed of three key components (Meyer et al., 2020):

(1) The NLP framework evaluates and analyses the user request and generates a natural language response. NLP is a branch of artificial intelligence that allows computers to communicate using natural human language. It studies how computers can understand and manipulate natural language in text or speech. The user request is evaluated and analyzed, generating a natural language answer (Srilekha et al., 2019). This usually consists of NLU and natural language generation (NLG) (E. Adamopoulou and Moussiades, 2020). NLU extracts context from unstructured human language input and responds by interpreting the user's purpose. Intent categorization and entity extraction are supported by NLU, which considers context (Jung, 2019). NLG, on the other hand, is the process of producing a human-like response using a data-driven approach (Samyn, 2019).
(2) The dialog manager compares user input to associated backend systems and extracts text or performs functions. This section controls and updates the context of the conversation.The dialog manager oversees matching ideas from the user's inquiry to knowledge domain rules. If the chatbot cannot gather the essential context information, it will ask the user for additional context information to fill in the gaps. After the purpose is identified, it also asks follow-up questions to promote user participation and minimize the chance for error (Abu Shawar, 2007; Hutapea, 2017).
(3) The backend comprises the essential application systems or databases. The chatbot uses external API (Application Programming Interface) calls or database requests to get the information from the backend to satisfy the user's goal. After the necessary data have been retrieved, it is sent to the Dialog Management NLP module. When rule-based chatbots are used, a KB is used. The KB must cover a wide range of user inquiries and have a variety of responses to the same user input to minimize the repetition of responses (Adamopoulou and Moussiades, 2020).

Recent research has found that simulated conversational agents that serve as educational companions or tutors can help students study more effectively (Hayashi, 2013; Stathakarou et al., 2020). Due to the use of Technology-Enhanced Learning (TEL) systems, where innovative technologies and artificial intelligence are used to allow greater flexibility, personalization, engagement, and motivation of learners, HE learning environments have evolved significantly in the last few decades (Hien, Cuong, Nam, Nhung, & Thang, 2018).

1.3. Chatbots in higher education settings

Driven by digitization, chatbots have been introduced in business contexts for customer support functions or to assist employees in their daily work and reduce service costs while handling multiple user inquiries at the same time independently of the availability of human resources. Consequent to positive experiences in the business context, the use of chatbots in academic settings has arisen as a new trend in natural language-based human-computer interaction research (Meyer von Wolff, Nörtemann, Hobert, & Schumann, 2020).

Rapid developments in education technology have resulted in integrating novel pedagogical techniques to improve the learning experience. In recent years, virtual learning support, E-learning, and M-learning have increased significantly (Bahja, Hammad, & Butt, 2020). As the number of individuals attending HE has grown, so has the number of students per teacher, reducing staff availability to attend to individual student requests (Hien et al., 2018). Chatbots have been one of the key focal points in fusing educational techniques with modern technology (Bahja et al., 2020).

Student engagement, specifically its lack of it, is currently challenging in HE (Studente & Ellis, 2020). Numerous universities have been using chatbots to help students with various activities, such as obtaining information on the degree of students' satisfaction in a given course (Wambsganss et al., 2020), administrative support (Sweidan et al., 2021), supporting enrolment (Konecki et al., 2015), but also in more academic tasks, such as, simulating a patient (Anubharath et al., 2019), collaborative learning (Michos et al., 2020) and writing support (Resch & Yankova, 2019). Chatbots as educational agents have been designed for a range of purposes, such as tutoring (Aguilar-Mejía & Tejeda, 2020), language learning (Pham et al., 2018), teachable agents (Ceha et al., 2021), and virtual patients (Halan et al., 2014).

When applied to HE, chatbots can provide individual assistance to students and improve interaction, sociability, and information acquisition by changing the educational flow to be more interactive and dynamic while enabling communication with the system using natural language in an intuitive and user-friendly way (Ondáš et al., 2019). Learning participation—crucial for engagement and

satisfaction factor—is paradoxically more attainable via a chatbot experience since the learner can choose whether and when to react (Studente & Ellis, 2020). Applying a chatbot in educational contexts provides learners with rapid access to information without time-demanding searching (Ondáš et al., 2019; Sjöström et al., 2018).

As a result of the covid 19 outbreak that affected 98.5% of students, the use of chatbots in education has increased exponentially (Bahja et al., 2020).

1.4. Objectives

We present a scoping literature review of chatbots in HE in this work. Previously untapped areas of literature are investigated to map the research and identify gaps in knowledge. The quality of the chatbots presented in the research was also assessed using the usability criteria defined in the International Organization for Standards (ISO) 9241–11 guidelines.

1.4.1. Research questions

The present review aims to answer the following questions:

RQ1. With what purpose are chatbots being used in HE settings?

RQ2. What are chatbots' main benefits and challenges when applied to HE settings, and what metrics are used to evaluate the interaction?

RQ3. Is the feasibility of the chatbots assessed?

RQ4. Is the quality of the chatbots generally evaluated?

2. Methods

A scoping review of empirical studies and public sources was conducted to identify literature about the usage of conversational agents in higher education settings. Scoping reviews are different from other types of reviews since they aim to provide an overview or map of the evidence rather than a critically assessed and synthesized result/answer to a specific topic (Munn et al., 2018). Also, they are an excellent technique for determining the extent of coverage of a body of literature on a specific issue, as they clearly indicate the volume of literature and studies available. These help analyze emerging data when it's still unclear what other, more specific issues a more precise systematic review could answer (Munn et al., 2018). The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Extension for Scoping Reviews standards were followed (Tricco et al., 2018).

As proposed by Munn et al. (2018), the PCC mnemonic (Population, Concept, and Context) was used to clarify the primary questions and aims of the current review. Any significant stakeholder in higher education settings, such as students and faculty, was described as "population". "Use of Chatbots" was defined as the concept, with "higher education" as the context. These terms, as well as their synonyms, were included in the searches.

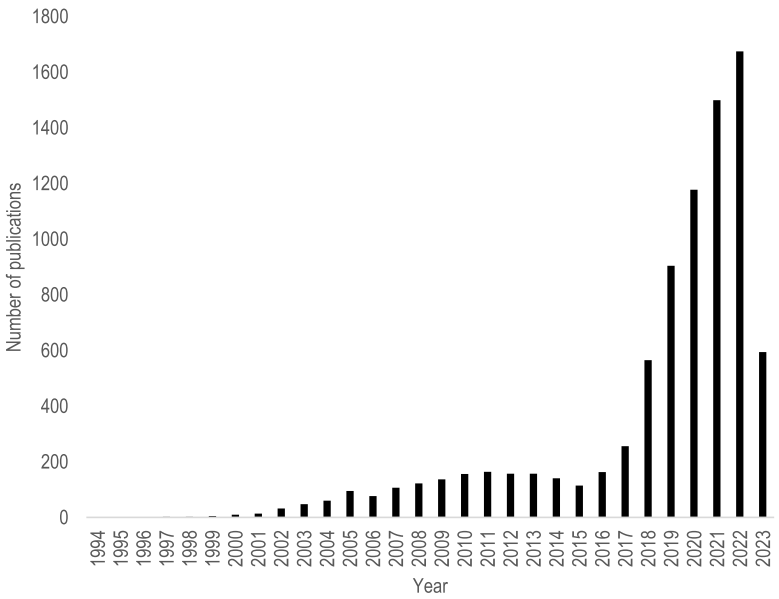

Fig. 2. Number of publications per year on Scopus ® from 1994 to May 2023. Search query: Chatbots OR Conversational Agents.

A literature search within the PubMed®, Web of Science – Core Collection®, and SCOPUS-ELSEVIER® electronic databases was conducted between January and May 2023. Since a quick review of the SCOPUS-ELSEVIER® electronic databases revealed an exponential increase in publications about chatbots throughout this time, documents were collected between 2010 and 2023 (Fig. 2).

Chatbots are referred to in the literature using various terms. However, the core premise is the same, whether you name them digital assistants, conversational interfaces, or just chatbots: achieve a goal by interacting with a machine in a dialogic manner using natural language. Hence, the search used terms used as synonyms present in literature combined into the following query:

((chatbot[Title/Abstract]) OR (Conversational agent[Title/Abstract])) OR (educational bot[Title/Abstract])) OR (pedagogical comversational agent[Title/Abstract])) OR (intelligent assistant[Title/Abstract])) OR (dialog system [Title/Abstract])) OR (teacherbot[Title/Abstract])) OR (computer based conversational agent [Title/Abstract])) OR (chatterbots [Title/Abstract])) AND (*education[Title/Abstract]))) OR (technology mediated learning [Title/Abstract])) OR (e-learning[Title/Abstract])) OR (computer supported collaborative learning [Title/Abstract]

From 2010 through May 2023. A total of 2692 publications were extracted, of which 500 were eliminated due to duplication of results.2192 were considered for inclusion and advanced to the screening phase. Studies identified through database searches were uploaded to Rayyan Management Software (Ouzzani et al., 2016). The software was used to conduct study selection, data records, search results, and eligibility criteria.

2.1. Selection criteria

The screening was conducted in two stages: initial screening and full-text screening. In both stages, articles were examined based on the inclusion and exclusion criteria. Initial screening was conducted by three authors, who examined articles based on title and abstract.

2.1.1. Initial screening

"This review considered studies related to the use of chatbots in HE settings. The following study designs met the inclusion criteria:

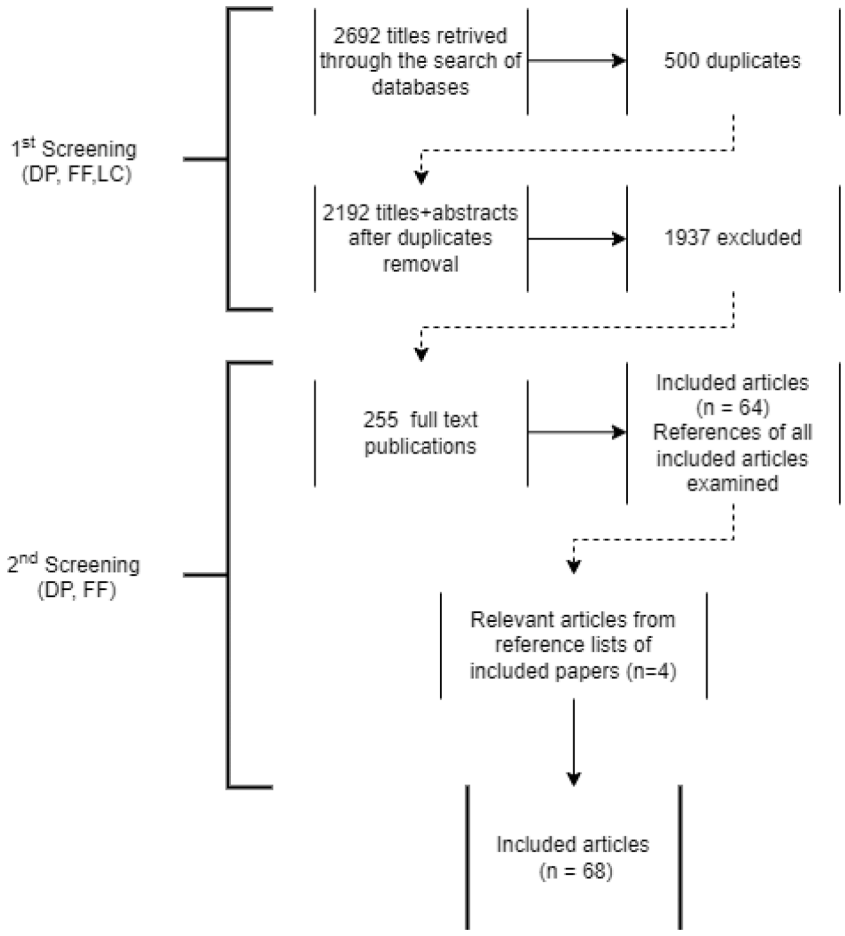

Fig. 3. Process used for study search and selection.

(i) use of chatbots in HE settings
(ii) studies written in English.

Studies were eliminated when:

(i) provided a systematic review or meta-analysis;
(ii) reported the building of the chatbot or design guidelines without a study of the population feedback on that instrument
(iii) were not written in English
(iv) presented incomplete data
(v) limited to technically improvement of chatbot
(vi) were exclusively theoretical

A final decision regarding the selected studies was made through discussion.

2.1.2. Full text screening

Initial Screening resulted in 255 publications eligible for the second stage of screening. In the second stage, one reviewer independently conducted a full-text screening with the papers approved in the first stage. In case of doubt, the reviewer discussed this decision with the other authors to reach a final decision. A final set of 64 publications was included in this review. By reference checking, another four publications were added to the final list. The process used for study search and selection is detailed in Fig. 3.

2.2. Data charting process

After the screening of the selected articles based on the predefined inclusion and exclusion criteria, the authors performed a thorough extraction process to synthesize relevant information. Any disagreements or discrepancies during the extraction process were resolved through consensus. The articles were categorized based on various parameters, including title, year, country of origin, population characteristics (gender, age), setting/purpose of the chatbot, domain of the chatbot, language, platform, type of publication, feasibility, metrics, and key outcomes.

The authors followed a structured data extraction form or template throughout the extraction process to ensure consistency and standardization. This form facilitated the systematic recording of relevant information from each article.

2.3. Data synthesis

The extracted data was synthesized using a narrative approach. The features of the included studies were assessed using descriptive statistics (frequency).

2.4. Quality assessment of conversational systems

Until a framework has been developed and broad consensus on the assessment criteria established, chatbot developers are bound to rely on the framework offered by International Organization for Standards (ISO) 9241–11 (International Organization for Standardisation, 2018) for assessing usability due to the lack of a shared assessment methodology specific to these systems (Al-Ghadhban & Al-Twairesh, 2020; Federici et al., 2020). Although these criteria have not yet been modified to meet the special needs of chatbots, usability is a critical aspect to consider to perceive the quality of engagement (Federici et al., 2020). ISO 9241 describes usability as: “software is usable when it allows the user to execute his task effectively, efficiently, and with satisfaction in the specified context of use.” (Al-Ghadhban & Al-Twairesh, 2020; International Organization for Standardisation, 2018). Effectiveness, efficiency, and satisfaction are the three most crucial usability traits in a specific use context. Effectiveness relates to the system’s ability to achieve the user’s objectives. Efficiency relates to how effectively resources are used to meet the user’s objectives. The term "satisfaction" relates to how happy users are with the system (Al-Ghadhban & Al-Twairesh, 2020). The framework provided by the ISO 9241–11 (Bevan et al., 2015) was used to assess if the selected papers studied the three strands mentioned of usability described in the ISO: Effectiveness; Efficiency and Satisfaction. Table 1 describes the measures of effectiveness, efficiency, and satisfaction measures considered in the present work.

Table 1
Measures of Effectiveness, efficiency and satisfaction. (Adapted from Bevan et al., 2015).

Effectiveness	Efficiency	Satisfaction
Tasks Completed	Task Time	Overall Satisfaction
Objectives Achieved	Time Efficiency Cost-Effectiveness	
Errors in a Task	Fatigue	

Table 2
Characteristics of selected papers.

Refs.	Year	Country	Language of chatbot	Platform	Type of publication
Abbas et al. (2022)	2022	UK	English	Differ	Original Paper
Abutaleb & Yun (2022)	2022	UK	English	DialogFlow	Conference Proceeding
Aguilar-Mejía & Tejeda (2020)	2020	Mexico	Spanish	NR	Conference Proceeding
Al Kahf et al. (2023)	2023	France	French	NR	Original Paper
Al-Abdullatif et al. (2023)	2023	Saudi Arabia	NR	WhatsApp	Original Paper
Alqaidi, Alharbi, & Almatrafi (2021)	2021	Saudi Arabia	English	DialogFlow	Conference Proceeding
Amini, Boustani, & Lisetti (2021)	2021	USA	English	NR	Original Paper
Anubharath et al. (2019)	2019	Singapore	English	NR	Conference Proceeding
Anumala, Chintalapudi, & Yalamati (2022)	2022	India	English	IBMWatson	Original Paper
Atmosukarto et al. (2021)	2021	Singapore	NR	NR	Original Paper
Campillos-Llanos et al. (2021)	2021	France	English, French, Spanish	NR	Original Paper
Carayannopoulos (2018)	2017	Canada	English	NR	Conference Proceeding
Carreira, Silva, Mendes, & Oliveira (2022)	2022	Portugal	Portuguese	Rasa	Original Paper
Ceha et al. (2021)	2021	Canada	English	NR	Conference Proceeding
Chakraborty, Mishra, Kumar, Singh, & Hani (2022)	2022	India	NR	NR	Conference Proceeding
Chang, Hwang, & Gau (2022)	2020	Taiwan	Mandarin	NR	Original paper
Chang et al. (2022)	2022	Taiwan	English	NR	Original paper
Chaudhry, Sarwary, El Refae, & Chabchoub (2023)	2023	United Arab Emirates	English	ChatGPT	Original Paper
Chen, Cheng, & Heh (2021)	2021	Taiwan	Mandarin	Line	Conference Proceeding
Clarizia et al. (2018)	2018	Italy	English	NR	Original paper
Co, Yuen, & Cheung (2022)	2022	Hong Kong	English	DialogFlow	Original Paper
Dupuy et al. (2021)	2020	France	French	Unity technologies	Original paper
Durak (2023)	2022	Turkey	English	FlowXO and Slack	Original Paper
El Hefny et al. (2021)	2021	Egypt	English and Arab	Facebook messenger	Book Chapter
Griol, García-Herrero, & Molina (2011)	2011	Spain	Spanish	NR	Conference Proceeding
Gupta & Chen (2022)	2022	USA	English	NR	Original Paper
Kumar, Yu, Chung, Shi, & Williams (2023)	2023	Canada	English	NR	Conference Proceeding
Hsu, Chan, & Yu (2023)	2020	Taiwan	English	NR	Original Paper
Halan et al. (2014)	2014	USA	English	Virtual Patient Pipeline	Conference Proceeding
Halan, Sia, Crary, & Lok (2015)	2015	USA	English	Virtual Patient Pipeline	Conference Proceeding
Han et al. (2022)	2022	South Korea	English	NR	Original Paper
Hayashi (2013)	2013	Japan	English	Java	Conference Proceeding
Hobert (2019)	2019	Germany	English	NR	Conference Proceeding
Kumar (2021)	2021	Malaysia	English	Telegram	Original Paper
Khalil & Rambech (2022)	2022	Norway	Norwegian	Telegram	Original paper
Konecki et al. (2015)	2015	Croatia	English	NR	Conference Proceeding
Kong et al. (2021)	2020	Singapore	NR	NR	Original Paper
Krassmann et al. (2018)	2018	Brazil	Portuguese Br	AIML	Original paper
Lee et al. (2020)	2020	China	English	DialogFlow	Conference Proceeding
Meshram, Naik, VR, More, & Kharche (2021)	2021	India	English	Rasa	Conference Proceeding
Tanana, Soma, Srikumar, & Atkins (2019)	2019	USA	English	NR	Original Paper
Mokmin & Ibrahim (2021)	2021	Malaysia	English	DialogFlow and telegram and Whatsapp	Original Paper
Moldt, Festl-Wietek, Mamlouk, & Herrmann-Werner (2022)	2022	Germany	NR	NR	Original Paper
Moore et al. (2022)	2022	Australia	English	NR	Original Paper
Neo et al. (2022)	2022	Malaysia	English	NR	Original Paper
Pham et al. (2018)	2021	Vietnam	Vietnamese	Rasa	Original Paper

(*continued on next page*)

Table 2 (*continued*)

Refs.	Year	Country	Language of chatbot	Platform	Type of publication
Okonkwo & Ade-Ibijola (2021)	2021	South Africa	English	SnatchBot	Original Paper
Ondáš et al. (2019)	2019	Slovakia	Slovak	DialogFlow	Conference Proceeding
Perez-Mercado et al. (2023)	2023	Spain	English	Laravel	Original Paper
Resch & Yankova (2019)	2019	Germany	English	Telegram messenger	Original Paper
Ryong, Lee, & Lee (2023)	2023	South Korea	English	NR	Original Paper
Saiz-Manzanares et al. (2023)	2023	Spain	Spanish	UBUVirtual Platform	Original Paper
Sevgi et al. (2023)	2023	Turkey	English	ChatGPT	Original Paper
Shim, Menkhoff, Teo, & Ong (2023)	2023	Singapore	NR	DialogFlow	Original Paper
Shorey et al. (2019)	2019	Singapore	English	Unity 3D and DialogFlow	Original Paper
Srinivasan et al. (2022)	2022	Sri Lanka	English, Tamil, and Thanglish	NR	Original Paper
Sweidan et al. (2021)	2021	Jordan	English and Arabic	Android Studio and JAVA	Conference Proceeding
Talan & Kalinkara (2023)	2023	Turkey	English	ChatGPT	Original Paper
Tan et al. (2021)	2021	Malaysia	NR	NR	Conference Proceeding
Tian, Risha, Ahmed, Lekshmi Narayanan, & Biehl (2021)	2021	USA	English	NR	Conference Proceeding
Valdivieso & Luzoz (2021)	2021	Spain	Spanish	NR	Original Paper
Vazquez-Cano, Mengual-Andres, & Lopez-Meneses (2021)	2021	Spain	Spanish	NR	Original Paper
Villegas-Ch et al. (2021)	2021	Ecuador	English	Web page, Facebook messenger, Whatsapp	Original Paper
Wambsganss et al. (2020)	2020	USA	English	NR	Conference Proceeding
Lee et al. (2022)	2022	Taiwan	Mandarin	T-BERT	Original Paper
Yeh, Chang, Ho, & Ma (2021)	2021	Taiwan	Mandarin	NR	Original Paper
Yi, Ray, & Segall (2023)	2023	USA	English	NR	Original Paper
Zaky (2023)	2023	Egypt	Arabic And English	WhatsAuto - Whatsapp	Original Paper

3. Results

3.1. Selection of sources of evidence

A total of 2692 citations were found while searching the previously specified internet databases. After removing 500 (18.6%) duplicate citations, 2192 (81.4%) titles and abstracts were assessed. The initial screening process resulted in 1937 titles and abstracts being rejected. Most of these were not included because the search key terms were so broad as to retrieve some articles that did not fit this paper's objectives and focus. The remaining articles were eliminated as they did not meet the inclusion criteria.

Full text screening resulted in 187 (8.5%) citations being eliminated (Fig. 3). Reference checking yielded a total of 4 studies. In total, 70 studies were considered in this research.

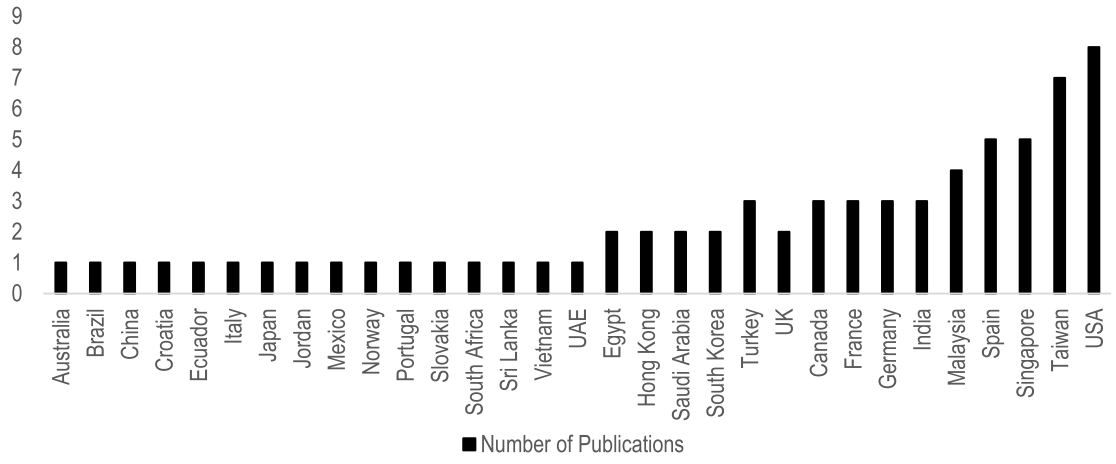

Fig. 4. Number of publications per Country.

3.2. Characteristics of selected papers

All of the studies were conducted between 2010 and 2023, as stated in Table 2; this was a deliberate decision to exclude studies completed before 2010 to bring any findings up to date. The countries of the studies are represented in Fig. 4. Most publications were original papers (n = 48), with English (n = 49) being the most common chatbot language. Although some studies did not specify which platform was used for the chatbot, WhatsApp, Telegram, Facebook Messenger and DialogFlow ® were among the most used.

Around 50% of the selected papers reported sample sizes, ranging from 28 to 215 participants. The average age of the participants ranged from 16 to 31 years old, with gender differences depending on the domain in which the conversational bots were used.

RQ1. With what purpose are chatbots being used in higher education settings?

Concerning RQ1, implementation objectives for chatbots in education were extracted. Most chatbots (78%) were employed for educational purposes, administrative support, or combination.

Regarding educational chatbots, the fields where most chatbots were used were health and computer science (n=20 and n=6, respectively), as well as learning process aid (n=8). The remaining educational chatbots were used in language learning, mathematics, physics, geology, writing support, chemistry and physics, and business (Fig. 5).

Most of the objectives for chatbots in education can be classified into one of four categories based on the analysis of the main goal of the selected publications: interpersonal skills training, logistic support, educational support, or other (Fig. 6).

Chatbots are reported to have been successfully used for interpersonal skills training, since they enable simulated interactions between healthcare students and virtual patient agents (Halan, Sia, Crary, & Lok, 2015). This is especially important in the pandemic situation since face-to-face teaching has been limited. In terms of logistic support, chatbots can often answer queries about course materials, such as course objectives and learning outcomes, class schedules and locations, professors' contact information, test dates and details, and due dates for various coursework (Lee et al., 2020). Chatbot systems are primarily used in the educational context for teaching and learning- educational support. Even though they were too different to group consistently, other objectives, such as collaborative learning, the effect of utilizing a humorous chatbot, writing support *etc*., make up a significant portion of these articles (18 %) (Fig. 6). When a publication indicated more than one goal was allocated to several categories.

The articles were divided into the following categories based on the role of the CA in education: when the chatbot's goal was to help the student with several tasks, it was defined as an assistant. When the chatbot answered questions on the course materials, it is referred to as a tutor. When the CA was a teachable agent, it was referred to as a student, and when the CA simulated a patient, it was referred to as a Virtual Patient (Table 3).

As shown in Fig. 7, most chatbots (55%) were classified as tutors, whereas the majority of health-related chatbots were classified as Virtual Patients (63,6%). Considering the remaining items, 74% of the systems had the role of tutor, and around 26% were assistants. Only one study researched the role of a chatbot as a student (2%).

RQ2. What are chatbots' main benefits and challenges when applied to HE settings, and what metrics are used to evaluate the interaction?

The feasibility, challenges, metrics, and key results for each included study is present in Table 4.

A challenge was a technical or population-related obstacle to the study's outcomes. Of all the publications, 42 described the challenges. 23.8% of studies described technical challenges such as need for design improvement (Aguilar-Mejía & Tejeda, 2020; Atmosukarto et al., 2021; Tian, Risha, Ahmed, Lekshmi Narayanan, & Biehl, 2021), NLP-related issues (Atmosukarto et al., 2021; Lee et al., 2022; Moore et al., 2022; Resch & Yankova, 2019; Villegas-Ch et al., 2021), and app-maintenance issues (Khalil & Rambech,

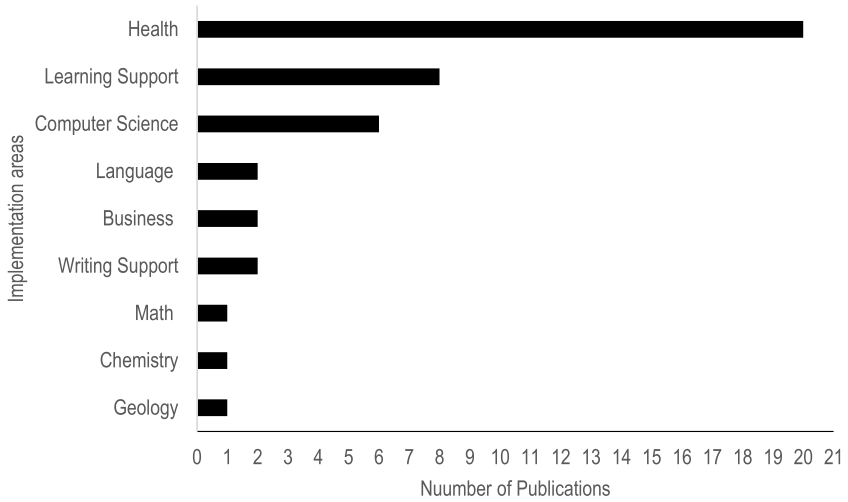

Fig. 5. Number of publications of educational Chatbots in different implementation areas.

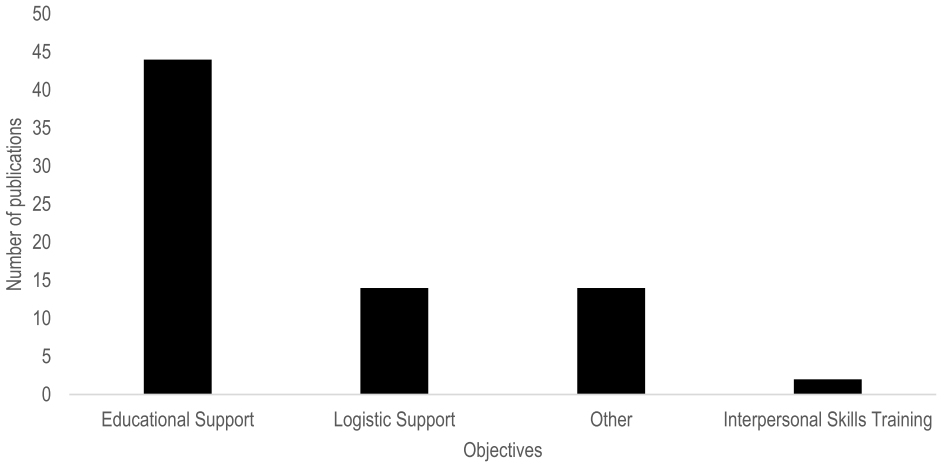

Fig. 6. Objectives for Chatbots in education based on the analysis of the selected publications.

2022). Another reported challenge was the lack of training data and/or limitations of a knowledge base (Anubharath et al., 2019; Chang et al., 2022; Krassmann et al., 2018; Resch & Yankova, 2019; Sevgi et al., 2023; Talan & Kalinkara, 2023; Zaky, 2023) and limited text comprehension (Ceha et al., 2021; Clarizia et al., 2018; Resch & Yankova, 2019). Regarding population-related challenges, around 17% of studies reported a limitation the population homogeneity. When studying the effect of a humorous chatbot on a population, Ceha et al. (2021) refer to reducing the generalization of results due to using a very homogeneous population while stating that the achieved results can differ for longer exposure. Limitations regarding novelty effects, fatigue and user expectations were found in five studies. Furthermore, the study, e.g Ceha et al. (2021), only focused on two types of linguistic humor, while many more exist. Other challenges relate to the fact that the chatbot was too obvious (Dupuy et al., 2021), user unfamiliarity (Krassmann et al., 2018), low number of participants (Chang et al., 2022; Durak, 2023; Lee et al., 2020; Shim, Menkhoff, Teo, & Ong, 2023; Tan et al., 2021), while five studies reported that the systems did not provide the necessary information's when asked (e.g.,Han et al., 2022).

The metrics used to evaluate the chatbots' performance and the results achieved by using chatbots were reported by 96% of the studies (n=65). Of these, 26 reported the use of user satisfaction questionnaires and 16 used validated scales, such as MSLQ (e.g., Al-Abdullatif et al., 2023), System usability scale (e.g., El Hefny, Mansy, Abdallah, & Abdennadher, 2021) and Learning strategies Scale (Saiz-Manzanares et al., 2023). Krassmann et al. (2018) also reviewed the system's log records. Two of the publications assessed the chatbot through knowledge gain. Aguilar-Mejía and Tejeda (2020) calculated Hake's g to understand the gain in the students' conceptual understanding, and Hayashi (2013) compared the pretest to posttest scores and calculated the gain. Both publications presented satisfactory results. Self-reported measures were also used to evaluate the performance of chatbots, mostly by using scales (Halan, Sia, Crary, & Lok, 2015), ratings of the user experience (Hobert, 2019) and self-reported response quality, syntactic readability and intensity of sentiments in the answers (Krassmann et al., 2018).

RQ3. Is the feasibility of the chatbots assessed?

According to Oxford's dictionary, feasibility is "the quality of being possible and likely to be achieved" (Oxford learner's Dictionaries). As a result, applying the technology in real-life situations was considered an indicator of feasibility. The selected publications were reviewed for feasibility studies according to the criteria (Table 4). In total, only about 12% of the studies referred to the feasibility of using chatbots in HE settings.

Dupuy et al. (2021) answered the recommendations of France's National Authority for Health: "never the first time with a patient" by designing and validating several chatbots in the last years in Bordeaux Medical School. This was also reported by Halan et al. (2014), who, by covering three separate years, established that health students could create a virtual patient in under five hours as part of coursework. In a particular work, a demo version of VPs was available online (Abutaleb & Yun, 2022). Halan, Sia, Crary, & Lok (2015) also conducted the study in a real-world educational setting, validating its feasibility and illustrating how virtual agent creation exercises can teach interpersonal skills in educational settings as part of course work. Anubharath et al. (2019) work is also being tested in a real-world setting, proving the approach's feasibility. Regarding administrative support, the system created by Villegas-Ch et al., 2021 managed to answer the questions of over 400 students, where 36 follow-up notifications were sent to the director. After 3 months, the language learning system developed by Pham et al., 2018 received around 50,000 queries.

RQ4. Is the quality of the chatbots generally evaluated?

The measures of effectiveness, efficiency and satisfaction considered in the present work are presented in Table 1. Although the main goal of this paper is not to understand the quality of chatbots *per se*, but to what extent the quality of chatbots is or is not evaluated in the different papers. The evaluation of papers is present in Table 5. Only one paper did not mention effectiveness. Around 70% of all publications reported user satisfaction, and 24 publications reported efficiency-related attributes. 26.5% of the publications reported

Table 3
Area of implementation and chatbot function (educational Purposes) based on the analysis of the selected publications.

Reference	Area of implementation	Chatbot function
Aguilar-Mejía & Tejeda (2020)	Education-Physics	Tutor
Anubharath et al. (2019)	Education-Health	Virtual Patient
Ceha et al. (2021)	Education-Geology	Student
Clarizia et al. (2018)	Education-Computer Science	Assistant
Dupuy et al. (2021)	Education-Health	Virtual Patient
Griol et al. (2011)	Education-Computer Science	Tutor
Halan et al. (2014)	Education-Health	Virtual Patient
Halan, Sia, Crary, & Lok (2015)	Education-Health	Virtual Patient
Hayashi (2013)	Education-Health	Assistant
Konecki et al. (2015)	Education-Computer Science	Tutor
Lee et al. (2020)	Education-Computer Science	Tutor
Resch & Yankova (2019)	Education	Assistant
Sweidan et al. (2021)	Education	Assistant
Chen et al. (2021)	Education	Tutor
Neo (2022)	Education	Tutor
Khalil & Rambech (2022)	Education	Tutor
Chang et al. (2022)	Education-Health	Tutor
Vazquez-Cano et al. (2021)	Education-Language Learning	Tutor
Okonkwo & Ade-Ibijola (2021)	Education-Computer Science	Tutor
Tan et al. (2021)	Education-Math	Tutor
Perez-Mercado et al. (2023)	Education-Computer Science	Tutor
Han et al. (2022)	Education-Health	Tutor
Co et al. (2022)	Education-Health	Virtual Patient
Chang et al. (2022)	Education-Health	Tutor
Valdivieso & Luzoz (2021)	Education	Assistant
Lee et al. (2022)	Learning Process Aid	Assistant
Neo et al. (2022)	Learning Process Aid	Assistant
Moore et al. (2022)	Education-Health	Virtual Patient
Campillos-Llanos et al. (2021)	Education-Health	Virtual Patient
Yeh et al. (2021)	Education-Health	Virtual Patient
Shim, Menkhoff, Teo, & Ong (2023)	Education-Business	Tutor
Al Kahf et al. (2023)	Education-Health	Virtual Patient
Hsu, Chan, & Yu (2023)	Education-Health	Tutor
Gupta & Chen (2022)	Education	Tutor/ Assistant
Srinivasan et al. (2022)	Education	Tutor/ Assistant
Carreira et al. (2022)	Education-Computer Science	Tutor
Abutaleb & Yun (2022)	Education-Health	Virtual Patient
Tian, Risha, Ahmed, Lekshmi Narayanan, & Biehl (2021)	Education	Tutor
Atmosukarto et al. (2021)	Education	Tutor
Amini, Boustani, & Lisetti (2021)	Education-Health	Virtual Patient
Kong et al. (2021)	Education-Health	Virtual Patient
Shorey et al. (2019)	Education-Health	Virtual Patient
Kumar (2021)	Education-Health	Tutor
MJ et al. (2019)	Education-Health	Virtual Patient
Mokmin & Ibrahim (2021)	Education-Health	Tutor
Saiz-Manzanares et al. (2023)	Education	Tutor
Hobert (2019)	Education	Tutor
Chen, Vicki Widarso, & Sutrisno (2020)	Education	Tutor
Sevgi et al. (2023)	Education-Health	Assistant
Talan & Kalinkara (2023)	Education-Health	Assistant

all three measures. All the papers, however, mentioned at least one of the criteria.

4. Discussion

Chatbots are being used in various industries, including education, although their application in real-life educational settings was, until recently, mostly nonexistent. The current study provides an overview of prior research on the application of chatbots in HE, including details on existing studies, benefits, and challenges. The present review also explored if the quality of chatbots is being evaluated in the literature. The main findings of the study are summarized and discussed in this section. To begin with, chatbots are commonly used in education in several capacities, including as service and teaching assistants. Their effectiveness as a service support tool and the success of education-oriented chatbots are extensively documented. Chatbots are the clearest examples of learning assistance, whose function is to help the teacher or repeat repetitious activities.

Concerning RQ1, the current study outlined the implementation goals, dividing the chatbots according to educational purposes, administrative support, or combination. This allowed the identification of the main impact areas of chatbots in HE. The clearest cases of using chatbots were in health and learning support. Further analysis allowed the categorization of chatbots in education, divided

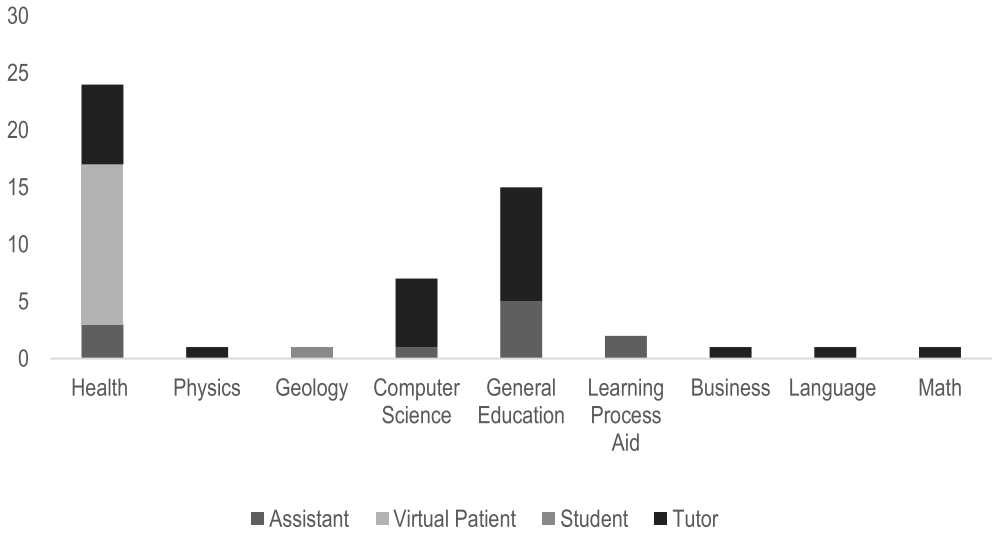

Fig. 7. Parallel between chatbot function and subjects of implementation.

into interpersonal skills training, logistical support, educational support, or other.

Chatbots are regularly used in education in different roles. The review outlined the four major roles described in the literature: assistant, tutor; student, and virtual patient. Chatbots have a demonstrated track record of success despite their flaws and issues. These systems can be extremely useful. However, few studies on the long-term effects of using chatbots in HE. This review also investigated the main challenges reported and the metrics used to evaluate the user experience. Addressing the challenges regarding the use of chatbots in HE exposes important factors that must be considered when developing a chatbot. As expected, the main challenges reported were technical or population related. Technical limitations, which include issues with NLP e Knowledge base limitations, emerge as a major obstacle. Even though there are other significant limitations outside just technological ones, developers must solve these issues to enhance user experience and chatbot acceptance. Even though there are other significant limitations outside just technological ones, developers must solve these issues to enhance user experience and chatbot acceptance. Even though there are other significant limitations outside just technological ones, developers must solve these issues to enhance user experience and chatbot acceptance and satisfaction.

Population-related issues can considerably impact the generalizability and application of the research findings. The studies covered in the review largely focus on homogenous user groups or educational settings, limiting the understanding of chatbot deployment across varied contexts. Users' early interest in and interactions with chatbots may differ from their long-term involvement, which can also impact how satisfied users are. Another described challenge was the small number of participants, since the research might not fully capture the whole spectrum of user experiences and opinions, which may restrict the robustness of the results. It is crucial to acknowledge and address these population-related challenges, to avoid overgeneralizing the findings and to motivate further investigation into these topics.

Overall, the results of the selected studies were positive and expected at the outset. Challenges sometimes avoid achieving the expected results (Anubharath et al., 2019; Resch & Yankova, 2019). It can be argued that the results of these publications prove the versatility and the promising aspects of this type of support system for university education.

The metrics employed in the publications offer insightful data on the efficiency and significance of these systems. Most users' responses and evaluations were highly subjective regarding reporting metrics. The present review revealed that the primary tool that metrics were reported in the selected publications was student questionnaires, with log records, knowledge gain and self-reported measures being the exception. Several publications used validated scales and questionnaires to gage users' experience, strengthening the reliability of the findings. Additionally, using Likert scales made it possible to understand user attitudes toward chatbot interactions and capture complex responses. It should be highlighted, however, that relying entirely on Likert scales and self-report measures may prove disadvantageous since is dependent on users' subjective assessments and may be affected by response biases. Unbiased information about the effectiveness and efficiency of chatbots was offered using analytical measures like performance metrics. Since there isn't a globally accepted criteria for measuring chatbot quality and user satisfaction, different research uses multiple assessment methods, making comparisons difficult. Despite this, this review provides readers with a thorough understanding of the metrics used to evaluate chatbots.

To our knowledge, there is no specific feasibility definition in the chatbot's context. The present study considered the feasibility of applying the technology in real-life situations. Very few articles reported using such systems in a real-life setting. This might mean students were not allowed to test the chatbot before its implementation. Allowing students to use chatbots prior to their formal adoption would provide the authors with more realistic and relevant input about the chatbots' strengths and issues, allowing for further enhancement and increased usage.

There is some disparity between techniques for evaluating the quality of a chatbot oriented to teaching, and techniques for

Table 4
Feasibility, metrics, challenges, and key results reported in selected publications.

Refs.	Setting	Feasibility	Metrics	Challenges	Key Results
Abbas et al. (2022)	Use of Differ to assist students with social integration	NR	Student Feedback	NR	59% students actively engaged with Differ, 63% found Differ useful and 59% would recommend Differ to a friend. Differ appropriate to boost student engagement among non-standard student cohorts.
Abutaleb & Yun (2022)	Create virtual patients in an OSCE setting	Demo version available online.	Questionnaire Student's performance	Low adherence of students and staff	83% of users reached a diagnosis with varying degrees of confidence, 92% of users reported to use the SP in the future.
Aguilar-Mejía & Tejeda (2020)	Learning Physics	NR	Hake's g	Design of the didactic sequence needs to be improved	There was an increase in students' conceptual understanding(Hake's g).
Al Kahf et al. (2023)	Impact of systems on students' success rate in their end-term exams	NR	Randomized Controlled Trial	weak adherence of students Single class with a limited number of students Single course Llimited number of clinical scenarios in pulmonology. Use of only multiple-choice questions on VP	Differences in scores on the pulmonology subtest over the academic year were significantly higher among Gamers and Users vs Controls. Trend to a better correlation when users were evaluated on a subject covered by Chatprogress.
Al-Abdullatif et al. (2023)	Effect on the motivation and learning strategies of postgraduate students	NR	Motivated strategies for learning questionnaire	Small Sample Size One subject Short Study duration	Chatbot system contributed to reducing distractions and cognitive load. Participants in the experimental group who used the chatbot system were more motivated to accomplish their learning tasks than those in the control group
Alqaidi et al. (2021)	Development of a chatbot for college students to communicate and exchange common experiences and interests.	NR	User satisfaction	NR	The systems acts as a supporting environment for students with overall satisfaction of 98%.
Amini, Boustani, & Lisetti (2021)	modeling rapport from a corpus of annotated videos of counseling sessions	NR	Heerink's questionnaire for user acceptance; Bartneck's Godspeed questionnaire for ECA features randomly controlled experiment	NR	Chatbot improved user's attitude,intention to use the system, perceived enjoyment, perceived sociability, perceived usefulness, social presence, and trust.
Anubharath et al. (2019)	Understand usability of the system	NR	Questionnaire	Lack of training data.	Chatbot model predicted incorrect responses that led participants to feel frustrated. A total of 2169 user interactions were performed with the chatbot.
Anumala et al. (2022)	QA chatbot	NR	NR	NR	Chatbot is helpful in guiding students with correct and relevant sources of information, especially for international applicants.
Atmosukarto et al. (2021)	E tutor- Chemistry	Chatbot scheduled for deployment in 2021 to assist 280 freshmen.	Questionnaire	Technical NLP Issues UX/Flow issues Unclear answers Missing content	Students prefer a chatbot that is available 24/7, providing to the flexibility

(*continued on next page*)

Table 4 (*continued*)

Refs.	Setting	Feasibility	Metrics	Challenges	Key Results
					of students' study schedule and preference.
Campillos-Llanos et al. (2021)	Design of a dialog-enabled VP system that can cope with a variety of clinical conditions.	35 different cases from 18 different specialities.	5-point Likert-scale Evaluation framework of dialog systems Dialog system's correctness	Technical issues, Lack of machine/deep learning approaches.	Mean user evaluation score was 4.06 out of 5. On average, around 74.3% of replies were correct, 14.9% incorrect, and in 10.7% the system behaved cautiously by deferring a reply. In the user evaluation, all aspects scored higher in the 29 unseen cases than in the 6 seen cases.
Carayannopoulos (2018)	Aid students transitioning into university	NR	Questionnaire	NR	Chatbots are an effective mean to reduce student transition challenges.
Carreira et al. (2022)	E tutor- programming	NR	Questionnaire	Low adherence to questionnaire Explanations of functionalities	Pyo facilitates student learning.
Ceha et al. (2021)	Effect of humorous teachable agent	NR	Questionnaire	Homogeneous Population Interaction was short The study focused on only two types of verbal humor Limited to text comprehension and the topic of rock classification.	Humorous CAs can both enhance and detract from a learner's experience and outcomes. Self-defeating humor increases effort but not enjoyment. Affective humor promotes motivation and effort while self-defeating humor enhances effort but not enjoyment.
Chakraborty et al. (2022)	QA chatbot	NR	Performance Measures	NR	The chatbot is capable of generating the required information in response to the user's question concerning Students, faculty and university.
Chang et al. (2022)	Teach nursing students to handle obstretic vaccination	NR	Pre-test/Post-test Self efficacy Interview	Study Limited to subject	Chatbot had strong influence in Self-efficacy. Students generally believed that the chatbot was able to promote self-efficacy as well as learning engagement and performances.
Chang et al. (2022)	a knowledge-based chatbot to overcome the limitations of conventional teaching in order to improve students' learning efficacy	NR	Questionnaire Pre-test/ post-test	Non- representative study Population Lack of data relating to students' learning processes, Small sample.	The chatbot system effectively enhanced students' academic performance, critical thinking, and learning satisfaction.
Chaudhry et al. (2023)	test the ChatGPT capability of solving a variety of assignments	NR	quasi-experimental design	Did not involve students	GPT's has the capability to respond effectively to an assortment of tools used for assessing students' learning outcomes at the undergraduate level. ChatGPT's capability of passing academic integrity thresholds.
Chen et al. (2020)	E tutor- Chinese	NR	TAM Pre-test/post-test Interview	NR	ChatBot significantly improved the students' learning achievement TAM model showed that perceived usefulness was the predictor of behavioral intention, whereas

(*continued on next page*)

Table 4 (*continued*)

Refs.	Setting	Feasibility	Metrics	Challenges	Key Results
					perceived ease of use was not.
Clarizia et al. (2018)	Prototype of e tutor	NR	Questionnaire	Identification of the real user needs.	Students find the Chatbot to be simple to use and understand. When compared to other Chatbots, students claim that it is easier to use and more effective.
Co et al. (2022)	Evaluate the feasibility and efficacy of bedside clinical teaching using chatbot app, based on students' performance on clinical history presentation.	All students were able to gather necessary clinical history from the chatbot.	Comparison between groups Median Likert scores of user friendliness, keyword identification, student-chatbot interaction, efficiency of learning and overall experience Questionnaire	Usage of pilot questionnaire	Blinded assessment of students' performance in clinical history taking were comparable between the conventional group and chatbot group $p > 0.05$.
Dupuy et al. (2021)	Empathy Training	Yes	Questionnaire	Empathic questions were too easy	The students managed to interact appropriately with the system, as overall they had good scores.
Durak (2023)	Effect of chatbots on visual design self-efficacy, engagement, satisfaction, and learner autonomy in the context of digital visual design education	NR	Satisfaction Scale Learner autonomy Scale Online engagment scale Self-efficacy scale	low number of participants Majority males	Frequency of use of chatbot did not positively affect any of the visual design self-efficacy, engagement, satisfaction, and learner autonomy or learning engagement
El Hefny et al. (2021)	QA- iniversity admissions	NR	System Usability Scale Chatbot Usability Questionnaire	NR	The mean SUS score is 88.5, while the mean CUQ score is 87.3.
Griol et al. (2011)	autonomous learning and self-assessment for e-learning	NR	Questionnaire	NR	Students give high marks to the ease with which they may collect the data needed to accomplish the entire set of stated objectives in the exercises for the subject, as well as the adequacy of the interaction rate during the dialog.
Gupta & Chen (2022)	Usage of chatbots to support inclusive learning	NR	Thematic analysis	Technical Limitations	Chatbots provide the opportunity to support disadvantaged students with varied learning styles.
Halan et al. (2014)	Creating virtual patients	Study went for three years, during which, students were able to create a virtual patient in under five hours	Rate the applicability of the virtual patient exercise based on: • Learning dysphagia information • Improving clinical interviewing skills • Valuable learning experience • Enjoyable	Fatigue of the students toward the end of semester.	During both user tests, participants who created a virtual patient increased the percentage of discoveries made between the first and second patients. This improvement, however, did not last until the third virtual patient interview.
Halan, Sia, Crary, & Lok (2015)	Virtual Patient of discordant race creation to increase empathy	experimental study was conducted in a real-world educational environment as part of a health professions course	Empathic Communication Coding System scale	NR	Students who created and interviewed virtual patients of the same race were significantly more empathetic than students who created virtual patients with a discordant race.
Han et al. (2022)	Development and evaluation of a chatbot to improve students EFM nursing skills	NR	Pre-test/Post-test	Chatbot provided feedback based on predicted responses, but did not address	When compared to the control group, the chatbot group did not exhibit any statistically significant differences in knowledge,

(*continued on next page*)

Table 4 (*continued*)

Refs.	Setting	Feasibility	Metrics	Challenges	Key Results
				specific questions asked by the students;	clinical reasoning ability, confidence, or feedback satisfaction. However, compared to the control group, its members' interest in education and self-directed learning was much higher.
Hayashi (2013)	E tutor- understanding and explanation of concepts	NR	Gain Scores pretest-posttest	NR	Participants performed better receiving audio prompts from the agent.
Hobert (2019)	E tutor- Programming	NR	Likert scale	NR	All design principles were evaluated by the students as useful and easy to use. Students rated the usefulness of DP3 adaptive learning path, DP4 Q&A component, and DP5 automatic feedback provision very positively.
Hsu, Chan, & Yu (2023)	E tutor- Medical Terminlogy	NR	Experimental study	Difficulty to control the frequency and duration of students' use of chatbot.	The results of the analysis suggest that there was no significant difference between the performance of the control group and the experimental group.
Konecki et al. (2015)	E tutor- programming	NR	Questionnaire	NR	System is perceived as a useful and interesting learning tool and aiding mean and a facilitator of more interesting way of learning, easier learning, greater motivation, and willingness of students to spend more time in learning of programming, preferring this approach opposed to classic.
Kong et al. (2021)	VP employed for second-year medical students.	NR	Questionnaire	NR	87% of participants agreed that using system helped in remembering the content and 69% reported that increased their confidence and competence in history-taking
Krassmann et al. (2018)	Understand Student's mood when interacting	NR	Questionnaire Analysis of the logs recorded by the CA	User unfamiliarity Limitations of the knowledge base.	When interacting with the system, students emotional states influence the quality of the conversation as well as the user's perception of the tool in terms of interest, utility, and satisfaction.
Kumar et al. (2023)	Investigate the potential of chatbots in supplementing existing mental well-being resources provided by universities.	NR	Likert Scale	NR	Students rated interaction with both the friend- LLM bot higher than the existing university chatbot.
Kumar (2021)	Efects of chatbot in team-based projects	NR	MSLQ and modified MSLQ Need for Cognition Scale–6 Creative Self-Efficacy Team Assessment Survey Questions	Users Expectations Ethical implications	Improved learning performance Facilitated collaboration among team Perception of learning, need for cognition, motivation, and creative self-efficacy were not influenced by system.
Lee et al. (2020)	Answer students Questions	NR	Questionnaire	Number of participants in the evaluation is low	Infobot is a helpful learning tool that allows for real-time problem solution and

(*continued on next page*)

Table 4 (*continued*)

Refs.	Setting	Feasibility	Metrics	Challenges	Key Results
				Limited information obtained in the focus group from these participants.	a quick review of course materials.
Lee et al. (2022)	Effects of using chatbot for after-class review on learning achievement, motivation, and attitude	NR	Learning Motivation Questionnaire Self efficacy	Technical Issues With NLP The experimental process was based on a short period Novelty effects	Experimental group and control group had different learning results because of the different review methods. The experimental group had better self-efficacy. The experimental group had a better learning attitude.
Meshram, Naik, VR, More, & Kharche (2021)	QA chatbot	NR	Performance measures	NR	Precision, Accuracy & F1 Score values of 0.628, 0.725 and 0.669, respectively.
Tanana et al. (2019)	Training new skills in psychotherapy	NR	Pre-test/Post-test	Very Specific Study Population	12.1% of users did not complete all phases of the interaction with the simulated client. There were no significant differences between groups regarding overall satisfaction.
Mokmin & Ibrahim (2021)	Evaluate the usefulness, performance, and technology acceptance of a chatbot developed to provide health literacy	NR	Unified Theory of Acceptance and Use of Technology 2	NR	73.3% of the respondents found that the chatbot can help understand several health issues and provide a good conversation. The chatbot contributed a low percentage of exit 37% of users exited the application
Moldt et al. (2022)	Evaluate the effectiveness of a chatbot in assessing the stress levels of students in everyday conversations.	NR	Perceived Stress Questionnaire	Limited to the german-speaking region Gender, age, response time and conversation duration not considered Interpretability of the quantitative results	Increased stress levels in 43.4% of medical students.
Moore et al. (2022)	Determine the feasibility of and barriers to verbal engagement with a virtual agent.	NR	System Usability Scale Free-text comments	Technical issues- NLP data collection limited to written forms and observation Novelty effect	Verbal interaction with a virtual agent is viable for training staff ID of 8 factors identified to be key to the experience of users in VR training application System usability was classified as “poor” on the usability scale.
Neo et al. (2022)	Students readiness to accept chatbot when learning online, and its impact in their learning process	NR	Questionnaire	NR	Students present positive attitudes towards using MERLIN to learn when they were learning independently online. MERLIN chatbot is capable of creating and supporting positive online learning experiences.
Pham et al. (2018)	E tutor- Language learning	After three months, the chatbot received around 50,000 questions.	Performance metrics	NR	Value of F1-score is 0.976, Accuracy is 0.971 and Precision is 0.979.

(*continued on next page*)

Table 4 (*continued*)

Refs.	Setting	Feasibility	Metrics	Challenges	Key Results
Okonkwo & Ade-Ibijola (2021)	E tutor- Python	NR	Questionnaire	NR	The design was user-friendly.Overall, learning Python programming was easier for students, improving their programming skills.
Ondáš et al. (2019)	Provide department information	NR	Questionnaire	NR	Students consider such services as attractive and helpful.
Perez-Mercado et al. (2023)	E tutor- SQL	NR	Questionnaire Number and quality of interactions	NR	ChatbotSQL provides personalised feedback and is effective in teacching SQL language.
Resch & Yankova (2019)	Writing support	NR	NR	Relevance of literature sources found Technical issues- NLP	The concept is found to be of help.
Ryong, Lee, & Lee (2023)	Influence of the chatbot's motivation strategies	NR	Pretest-post test Questionnaire	NR	Use of chatbot prior increases sense of enjoyment, the learner's recognitionof the goal and action plans, self-efficacy and intention to continue using.
Saiz-Manzanares et al. (2023)	Tutor- Self-regulated learning	NR	Learning strategies scale Metacognitive Scale Information Processing Support Scale IPSS	Very Specific Population Need to improve the feedback provided by the chatbot Need to improve how students' use of the chatbot is monitored throughout the semester	Level of degree being studied influenced the frequency of chatbot use and learning outcomes Levels of prior knowledge only influenced learning outcomes.
Sevgi et al. (2023)	How can chatGPT contribute to neurosurgery education by preparing case reports or questions, and its contributions when writing academic articles.	NR	NR	Questions and case examples were limited. More research is needed	While ChatGPT provided intriguing and interesting responses can't be considered a dependable source of information
Shim, Menkhoff, Teo, & Ong (2023)	Implementation of an experiential chatbot workshop integrated into an introductory undergraduate management course	NR	Proficiency scale Alignment with learner's expectation likert scale Intrinsic Motivation Engagement Chatbot-related competencies Workshop satisfaction	Covid 19 required replanning of workshop; Small sample size, limiting generalizations of results	Majority of respondents were pleased with the experiential learning chatbot 90.7% were satisfied with the experiential learning chatbot workshop; 81.4% of the respondents felt engaged and 81.3% of the participants reported moderate to high levels of competencies. 97.7% felt that the experiential chatbot workshop had met the expected learning outcomes.
Shorey et al. (2019)	Development of VPs for communicating skills development.	NR	User Acceptance Test	Content development Technological limitations Expectations management	Development of four case scenarios congruent with learning objetives.
Srinivasan et al. (2022)	Development of system to: A. Chatbots for University Administration and Students . B. Recommending solutions for psychological issues . C. Career Guidance Career y.	NR	Performance Measures Percieved Stress Scale	NR	Chatbot provides answers with high accuracy for 3 languages; Career guidance related questions and the answers were suitable. The psychological

(*continued on next page*)

Table 4 (*continued*)

Refs.	Setting	Feasibility	Metrics	Challenges	Key Results
	D. Performance prediction and learning materials recommendation				recommendation model was capable of identify students' psychological issues and propose solutions
Sweidan et al. (2021)	Covid19 support	NR	Questionnaire	NR	Feedbacks of most users were very positive.
Talan & Kalinkara (2023)	performance comparison of ChatGPT in an anatomy course with that of undergraduate students.	NR	Statistical Analysis	limited studies in the literature	ChatGPT's performance was better than the average performance of the students.
Tan et al. (2021)	E tutor- Math	NR	Likert scale Interview	Small Sample Study Limited to subject	Students' perception towards the use of conversational agent is favorable Positive qualitative feedback.
Tian, Risha, Ahmed, Lekshmi Narayanan, & Biehl (2021)	Developmet of learning assistant	NR	Questionnaire	Technical limitations Single session with users	ID of important design challenges
Valdivieso & Luzoz (2021)	E tutor- autonomous learning	NR	Likert scale TAM Model	NR	Increase in the percentage of students passing the subjects after using the chatbot tutoring system. TAM Scale results reveal that the tutoring system through the chatbot has a high utility .
Vazquez-Cano et al. (2021)	E tutor- Spanish	NR	Pre-test/Post-test	Non- representative study Population Personal variables were not taken into consideration.	The group of students who used the chatbot outside the face to face classroom, have substantially improved the results.
Villegas-Ch et al. (2021)	QA chatbot	400 individuals used the chatbot, 36 followup notifications were sent to the director.	Likert Scale	Technical Issues -NLP	Of the 36 followup, 2% became students of the degree. Chatbot has reduced the time spent on follow-up activities.
Wambsganss et al. (2020)	engage students to increase response quality of course evaluations	NR	Self-reported response quality by the user Syntactic readability based on the flesch-readability score Intensity of sentiments in the answers	Only a sample of course evaluation questions were asked. Novelty effects	Participants who used a CA expressed more satisfaction and were more likely to provide high-quality comments. The intention to use a CA for online course assessments is higher, according to the technology acceptance.
Yeh et al. (2021)	Chatbot to increase students' sensitivity and competencies in caring	NR	Peer Caring Measurement	Study performance was hard to evaluate	Successfully promoted the first-year students' sensitivity and behaviors in caring.
Yi et al. (2023)	Use of a Chatbot to for trainee recruitment	NR	Adapted System Usability Scale	Low adherence	84% reported that the chatbot provided a quick response, andhalf reported the friendliness. 22% indicated the chatbot was unable to help them, 11% reported strugles fining the informtion required.
Zaky (2023)	Chatbot to improve well-being and referencing skills	NR	Achievement Test: The test aimed to measure the cognitive achievement of APA referencing skills among postgraduate students.	Not enough background research Limited context	Statistically significant difference between the average scores of the experimental group in the pre and posttests. High level of digital well-being among the students

(*continued on next page*)

Table 4 (*continued*)

Refs.	Setting	Feasibility	Metrics	Challenges	Key Results
Khalil & Rambech (2022)	Helping students at the by making course content more accessible and immersive on mobile devices.	NR	Digital Well-Being Survey Agile development of the proposed chatbot Rambech Heuristics usability testing system Usability Scale	App Maintenance	Overall perception from students was positive. Students appreciated the efficiency of acquiring course content in a new novel format, both immersive and user friendly

evaluating the quality of a chatbot focused on technology. It is clear from the sample of publications chosen that there are still some differences between the various approaches to evaluating chatbots. Still, Radziwill & Benton (2017) concluded that evaluation procedures are often similar to the three characteristics of usability defined by the ISO appears to be valid when evaluating chatbots' quality.

4.1. Limitations

In this review, we set out to establish the potential of chatbots in education, and while we feel we accomplished the proposed goals, some limitations need to be addressed.

The first limitation regards the nomenclature given to chatbots. All too often in the literature, different names are given to the chatbots or/and different species of chatbots are given the same label and mixed in studies. This generates confusion and difficult interpretation of the results. Although we tried to cover this problem in our research design, we cannot be certain that this might have impacted the conclusions drawn in the present publication.

Secondly, the selected articles solely included journal and conference publications without considering other forms of publications.

Another significant limitation relates to the study populations. Generally, these are composed of students, leaving out the views of faculty staff -who play a major role in education. Furthermore, these studies' findings depend on population responses and willingness to participate, and their interpretation of the questions asked (especially in self-reported measures).

The last and most significant limitation is the application of categories, which could not be based on theoretical frameworks due to the novelty of the subject matter. This can potentially affect the interpretation of the data and conclusions drawn.

5. Conclusions and future research directions

The scoping review results highlight significant implications for chatbot design and integration in HE. To match chatbot implementations with the needs and preferences of keyplayers it is essential to adopt user-centric design methodologies and instructional models like ADDIE (Analysis, Design, Development, Implementation, Evaluation) or ASPIRE (Assess, Select, Prototype, Implement, Review, Enhance). While some studies investigated the specific requirements of users through methods such as interviews and questionnaires, it is crucial to ensure that this is a standard step in all implementation processes. Additionally, the study of UX and usability testing may produce visually appealing and intuitive chatbot interfaces that improve the user experience as a whole.Additionally, the study of UX and usability testing may produce visually appealing and intuitive chatbot interfaces that improve the user experience. The present work also emphasizes the importance of enhancing NLP models to guarantee that chatbots can efficiently understand and react to various difficult and subject-specific queries. With the introduction of hatGPT, this is a lot more attainable. Educational institutions may now enable chatbots to manage context, slang, domain-specific language, and multilingual dialogues utilizing powerful NLP and LLM like 15hatGPT, enabling more accurate and meaningful user interactions.

The study also emphasizes the significance of addressing the practical difficulties of adopting and sustaining chatbot solutions in HE. To guarantee the successful integration and sustainability of chatbot installations, institutions should assess their IT architecture, resource availability, and data security measures.

The review's insights into quality assessment techniques can help organizations set up efficient evaluation standards and ongoing improvement procedures. Chatbot replies can be improved with regular monitoring and user input gathering, which raises overall performance and user happiness. The review also emphasizes other crucial elements that are necessary for higher education institutions to use chatbots for better educational experiences effectively. These include user training and support, seamless integration of chatbots with current systems and support services while adhering to ethical considerations, and user training and support.

The study's findings provide a comprehensive overview of previous research on using chatbots in HE, including information on existing studies, advantages, and challenges. These publications show the versatility and the promising aspects of this type of support system for university education. However, despite the increased interest in chatbots, there is a lack of chatbot integration and evaluation into formal learning settings, which still needs to be done. Even though chatbots are often described as incredibly promising, Srivastava (2021) believes that the opportunity for this technology to reach its peak of significance has passed. Creating chatbots that can have natural and coherent conversations with humans has remained chiefly unfulfilled due to design and technical challenges, resulting in a gap between users' expectations and experience and subsequent chatbot abandonment (Srivastava, 2021; Wang et al.,

Table 5
Quality of chatbots according to ISO guidelines.

Refs.	Effectiveness	Efficiency	Satisfaction
Abbas et al. (2022)	Yes	–	Yes
Abutaleb & Yun (2022)	Yes	Yes	Yes
Aguilar-Mejía & Tejeda (2020)	Yes	–	–
Al Kahf et al. (2023)	Yes	–	Yes
Al-Abdullatif et al. (2023)	Yes	Yes	Yes
Alqaidi et al. (2021)	Yes	–	Yes
Amini, Boustani, & Lisetti (2021)	Yes	–	Yes
Anubharath et al. (2019)	Yes	Yes	Yes
Anumala et al. (2022)	Yes	–	–
Atmosukarto et al. (2021)	Yes	Yes	Yes
Campillos-Llanos et al. (2021)	Yes	–	Yes
Carayannopoulos (2018)	Yes	Yes	Yes
Carreira et al. (2022)	Yes	–	Yes
Ceha et al. (2021)	–	–	Yes
Chakraborty et al. (2022)	Yes	Yes	–
Chang et al. (2022)	Yes	–	Yes
Chang et al. (2022)	Yes	–	Yes
Chaudhry et al. (2023)	Yes	–	–
Chen et al. (2021)	Yes	–	–
Clarizia et al. (2018)	Yes	–	Yes
Co et al. (2022)	Yes	–	Yes
Dupuy et al. (2021)	Yes	Yes	Yes
Durak (2023)	Yes	Yes	Yes
El Hefny et al. (2021)	Yes	–	Yes
Griol et al. (2011)	Yes	Yes	Yes
Gupta & Chen (2022)	Yes	–	–
Kumar et al. (2023)	Yes	–	–
Halan et al. (2014)	Yes	–	Yes
Halan, Sia, Crary, & Lok (2015)	Yes	Yes	Yes
Han et al. (2022)	Yes	–	Yes
Hayashi (2013)	Yes	Yes	–
Hobert (2019)	Yes	–	–
Hsu, Chan, & Yu (2023)	Yes	–	Yes
Khalil & Rambech (2022)	Yes	–	Yes
Konecki et al. (2015)	Yes	Yes	Yes
Kong et al. (2021)	Yes	Yes	Yes
Krassmann et al. (2018)	Yes	Yes	Yes
Kumar (2021)	Yes	–	–
Lee et al. (2020)	Yes	–	Yes
Lee et al. (2022)	Yes	–	Yes
Meshram, Naik, VR, More, & Kharche (2021)	Yes	–	–
Mokmin & Ibrahim (2021)	Yes	Yes	Yes
Moldt et al. (2022)	Yes	Yes	Yes
Moore et al. (2022)	Yes	Yes	–
Neo et al. (2022)	Yes	–	Yes
Okonkwo & Ade-Ibijola (2021)	Yes	–	–
Ondáš et al. (2019)	Yes	Yes	Yes
Perez-Mercado et al. (2023)	Yes	–	–
Pham et al. (2018)	Yes	–	Yes
Resch & Yankova (2019)	Yes	Yes	Yes
Ryong, Lee, & Lee (2023)	Yes	–	Yes
Saiz-Manzanares et al. (2023)	Yes	–	Yes
Sevgi et al. (2023)	Yes	–	–
Shim, Menkhoff, Teo, & Ong (2023)	Yes	–	Yes
Shorey et al. (2019)	Yes	–	Yes
Srinivasan et al. (2022)	Yes	–	–
Sweidan et al. (2021)	Yes	Yes	Yes
Talan & Kalinkara (2023)	Yes	–	–
Tan et al. (2021)	Yes	–	Yes
Tanana et al. (2019)	Yes	Yes	Yes
Tian, Risha, Ahmed, Lekshmi Narayanan, & Biehl (2021)	Yes	Yes	Yes
Valdivieso & Luzoz (2021)	Yes	–	Yes
Vazquez-Cano et al. (2021)	Yes	–	–
Villegas-Ch et al. (2021)	Yes	Yes	–
Wambsganss et al. (2020)	Yes	Yes	–
Yeh et al. (2021)	Yes	–	–
Yi et al. (2023)	Yes	–	Yes
Zaky (2023)	Yes	–	Yes

2021).

Whilst studies have stressed the potential benefits of chatbots in education, issues such as technology limitations and the need for a natural language environment, among others, have made adoption difficult. Adopting advanced technologies like supervised learning also adds to the complexity of chatbot design, necessitating the usage of user-centric design techniques (Bahja et al., 2020). The directions for future research can be divided into two areas. The first takes a technological approach, where developers support creating and offering tools that allow teachers to integrate chatbots into their classes without difficulty. It is also important to improve the model's functionality when used in an educational context (Sjöström et al., 2018). NLP models must also be improved to avoid user fatigue (Ceha et al., 2021; Lee et al., 2020). The second is related to establishing chatbots' long-term effects on learning processes and outcomes.

One of the major concerns that researchers should consider in the future, which was not discussed in the scope of the present work but must be considered, is the ethical principles surrounding creating any AI-related system, i.e., chatbots. It's critical for scholars to look for legislative answers by establishing ethical rules and guidelines for employing chatbots in education (Gutiérrez et al., 2019). A new light was shed on this subject since the arrival of ChatGPT. Since it's launch, this innovative AI model has attracted a lot of interest for its versatility (Kung et al., 2023), reaching one million users in the first five days (Rudolph et al., 2023). But the question remains, what is ChatGPT's role, and how can it be disruptive to the world of education? In this work, we presented some publications that already demonstrate the use of ChatGPT in HE settings. However, there is still much to unveil regarding the impact of ChatGPT and other models with a similar design and the improved prospects offered by these systems. ChatGPT may encourage learners' autonomy and enhance their learning experience because it is customized to each learner's requirements and preferences. But, issues like ethics, data security, cheating, and plagiarism, must be researched before using ChatGPT – or any other bot- in HE settings.

*Structured practitioner notes**

What is already known about this topic

- As AI technology advances, educational shifts are taking place and chatbots are becoming increasingly integrated into our daily lives.
- Although they can be used in many different fields, they are especially important for the digital transformation of education.
- Recent studies have shown that virtual conversational agents act as study partners or instructors, and can improve students' academic performance.

What this paper adds

- The current paper summarises earlier studies on the use of chatbots in HE, together with information on their advantages and drawbacks.
- The current review also investigated whether chatbot quality is assessed in the literature. Chatbots are frequently employed in education in various capacities, such as service and teaching assistants.
- The review identified the four main roles discussed in the literature: virtual patient, student, tutor, and assistance. Despite its shortcomings and problems, chatbots have a proven track record of success.
- There is, however, little data available on the long-term implications of chatbot use in HE.

Implications for practice and/or policy

- Although studies have emphasized the potential advantages of chatbots in education, adoption has been challenging due to problems including technological constraints and the requirement for a natural language environment, among others. The complexity of chatbot design is increased using cutting-edge technology like supervised learning, demanding the application of user-centric design methodologies.
- There are two main areas where future research should focus. Firstly, developers assist in creating and providing technologies that make it simple for teachers to incorporate chatbots into their lessons. Enhancing the model's capability for usage in educational settings is also crucial. Improved NLP models are also necessary to prevent user fatigue. The second relates to assessing chatbots' usability in HE settings because additional empirical research is required to fully grasp their true potential.
- To determine the long-term impacts of chatbots on learning outcomes and processes, more research is needed. To determine the long-term impacts of chatbots on learning outcomes and processes, more research is needed.

CRediT authorship contribution statement

Daniela S.M. Pereira: Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Filipe Falcão:** Methodology, Writing – review & editing. **Lilian Costa:** Methodology, Writing – review & editing. **Brian S. Lunn:** Methodology, Project administration, Writing – review & editing. **José Miguel Pêgo:** Writing – original draft, Methodology, Project administration, Writing – review & editing. **Patrício Costa:** Writing – original draft, Methodology, Project administration, Writing – review & editing.

Declaration of Competing Interest

All authors declare no conflicts of interest.

Funding and acknowledgments

The authors acknowledge the Portuguese Foundation for Science and Technology (FCT) and Icognitus4All - It Solutions, LDA for providing a doctoral fellowship to DSMP (PD/BDE/150608/2020) through the PdDiHES program. This study was funded by National funds, through the FCT - project UIDB/50026/2020 and UIDP/50026/2020.

References

Abbas, N., Whitfield, J., Atwell, E., Bowman, H., Pickard, T., & Walker, A. (2022). Online chat and chatbots to enhance mature student engagement in higher education. *International Journal of Lifelong Education, 41*(3), 308–326. https://doi.org/10.1080/02601370.2022.2066213

Abu Shawar, B. (2007). Chatbots: Are they really useful? *LDV-Forum: Zeitschrift Für Computerlinguistik Und Sprachtechnologie, 22*(1), 29–49.

Abutaleb, A.-H., & Yun, B. (2022). In *Chameleon - A Framework for Developing Conversational Agents for Medical Training Purposes. International Conference on Autonomous Agents and Multi-Agent Systems 2022* (pp. 1887–1889). https://www.scopus.com/inward/record.uri?eid=2-s2.0-85134340057&partnerID=40&md5=32d7e8b67a891305a90c8e9d29fdace9.

Adamopoulou, E., & Moussiades, L. (2020a). In I. Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), *An overview of chatbot technology BT - Artificial intelligence applications and innovations* (pp. 373–383). Springer International Publishing.

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

Aguilar-Mejía, J. R., & Tejeda, S. (2020). Using virtual assistant for learning selected topics of physics. In *Proceedings of the IEEE integrated STEM education conference (ISEC)* (pp. 1–4). https://doi.org/10.1109/ISEC49744.2020.9397822

Al Kahf, S., Roux, B., Clerc, S., Bassehila, M., Lecomte, A., Moncomble, E., … Planquette, B. (2023). Chatbot-based serious games: A useful tool for training medical students? A randomized controlled trial. *PLoS One, 18*(3), Article e0278673. https://doi.org/10.1371/journal.pone.0278673

Al-Abdullatif, A. M., Al-Dokhny, A. A., & Drwish, A. M. (2023). Implementing the Bashayer chatbot in Saudi higher education: Measuring the influence on students' motivation and learning strategies. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1129070

Al-Ghadhban, D., & Al-Twairesh, N. (2020). Nabiha: An Arabic dialect chatbot. *International Journal of Advanced Computer Science and Applications, 11*(3), 452–459. https://doi.org/10.14569/ijacsa.2020.0110357

Alqaidi, S., Alharbi, W., & Almatrafi, O. (2021). A support system for formal college students: A case study of a community-based app augmented with a chatbot. In *2021 19th International Conference on Information Technology Based Higher Education and Training, ITHET 2021*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ITHET50392.2021.9759796.

Amini, R., Boustani, M., & Lisetti, C. (2021). Modeling Rapport for Conversations About Health with Autonomous Avatars from Video Corpus of Clinician-Client Therapy Sessions. In *, 12777. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 181–200). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-77817-0_15.

Anubharath, P., Chui, Y. P., Sng, J., Zhu, L., Tham, K., & Lee, E. J. D. (2019). Usability and user experience evaluation of virtual integrated patient. In *Proceedings of the ASCILITE 2019 - conference proceedings - 36th international conference of innovation, practice and research in the use of educational technologies in tertiary education: personalised learning. diverse goals* (pp. 272–281). One Heart.

Sandu, N., & Gide, E. (2019). Adoption of AI-Chatbots to Enhance Student Learning Experience in Higher Education in India. *18th International Conference on Information Technology Based Higher Education and Training (ITHET)* (pp. 1–5). Magdeburg, Germany. https://doi.org/10.1109/ITHET46829.2019.8937382

Anumala, R. R., Chintalapudi, S. L., & Yalamati, S. (2022). Execution of College Enquiry Chatbot using IBM virtual Assistant. *Proceedings - 2022 International Conference on Computing, Communication and Power Technology* (pp. 242–245). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/IC3P52835.2022.00058

Atmosukarto, I., Sin, C. W., Iyer, P., Tong, N. H., & Peng Yu, K. W. (2021). Enhancing Adaptive Online Chemistry Course with AI-Chatbot. In *TALE 2021 - IEEE International Conference on Engineering, Technology and Education, Proceedings* (pp. 838–843). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/TALE52509.2021.9678528.

Bahja, M., Hammad, R., & Butt, G. (2020). A User-Centric Framework for Educational Chatbots Design and Development. In *, 12424. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 32–43). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-60117-1_3.

Bevan, N., Carter, J., Harker, S., & Kurosu, M. (2015). ISO 9241-11 revised: What have we learnt about usability since 1998?. *Human-computer interaction: design and evaluation* (pp. 143–151). Springer International Publishing.

Brown, M., Mccormack, M., Reeves, J., Brooks, D. C., Grajek, S., Bali, M., & Horizon, E. (2020). *2020 EDUCAUSE Horizon Report TM Teaching and Learning*.

Campillos-Llanos, L., Thomas, C., Bilinski, É., Neuraz, A., Rosset, S., & Zweigenbaum, P. (2021). Lessons learned from the usability evaluation of a simulated patient dialogue system. *Journal of Medical Systems, 45*(7), 69. https://doi.org/10.1007/s10916-021-01737-4

Campillos-Llanos, L., Thomas, C., Bilinski, É., Zweigenbaum, P., & Rosset, S. (2019). Terminological and language resources for developing a virtual patient dialogue system in Spanish. *Procesamiento Del Lenguaje Natural, 63*(0), 205–208. http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/2019-63-31.

Carayannopoulos, S. (2018). Using chatbots to aid transition. *International Journal of Information and Learning Technology, 35*(2), 118–129. https://doi.org/10.1108/IJILT-10-2017-0097

Carreira, G., Silva, L., Mendes, A. J., & Oliveira, H. G. (2022). Pyo, a Chatbot Assistant for Introductory Programming Students. In *SIIE 2022*. In *24th International Symposium on Computers in Education*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/SIIE56031.2022.9982349.

Ceha, J., Lee, K. J., Nilsen, E., Goh, J., & Law, E. (2021). Can a humorous conversational agent enhance learning experience and outcomes?. In. In *Proceedings of the CHI conference on human factors in computing systems*. Association for Computing Machinery. https://doi.org/10.1145/3411764.3445068.

Chakraborty, D., Mishra, A., Kumar, V., Singh, S., & Hani, H. (2022). VOID - An Assistant (Chatbot) to Education Queries System. In *ASSIC 2022*. In *Proceedings: International Conference on Advancements in Smart, Secure and Intelligent Computing*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ASSIC55218.2022.10088400.

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

Chang, C. Y., Kuo, S. Y., & Hwang, G. H. (2022). Chatbot-facilitated nursing education: incorporating a knowledge-based chatbot system into a nursing training program. *Educational Technology & Society, 25*(1), 15–27.

Chaudhry, I. S., Sarwary, S. A. M., El Refae, G. A., & Chabchoub, H. (2023). Time to revisit existing student's performance evaluation approach in higher education sector in a new era of ChatGPT - A case study. *Cogent Education*, (1), 10. https://doi.org/10.1080/2331186X.2023.2210461

Chen, H. L., Vicki Widarso, G., & Sutrisno, H. (2020). A chatbot for learning Chinese: Learning achievement and technology acceptance. *Journal of Educational Computing Research, 58*(6), 1161–1189. https://doi.org/10.1177/0735633120929622

Chen, L. E., Cheng, S. Y., & Heh, J. S. (2021). Chatbot: A question answering system for student. In *Proceedings of the IEEE 21st international conference on advanced learning technologies, ICALT 2021* (pp. 345–346). https://doi.org/10.1109/ICALT52272.2021.00110

Clarizia, F., Colace, F., Lombardi, M., Pascale, F., Santaniello, D., Castiglione, A., Pop, F., Ficco, M., & Palmieri, F. (2018). *Chatbot: An education support system for student BT - Cyberspace safety and security* (pp. 291–302). Springer International Publishing.

Co, M., Yuen, T. H. J., & Cheung, H. H. (2022). Using clinical history taking chatbot mobile app for clinical bedside teachings-A prospective case control study. *HELIYON, 8*(6). https://doi.org/10.1016/j.heliyon.2022.e09751

Dolianiti, F., Tsoupouroglou, I., Antoniou, P., Konstantinidis, S., Anastasiades, S., Bamidis, P., Frasson, C., Bamidis, P., & Vlamos, P. (2020). *Chatbots in healthcare curricula: The case of a conversational virtual patient BT - Brain function assessment in learning* (pp. 137–147). Springer International Publishing.

Dupuy, L., de Sevin, E., Cassoudesalle, H., Ballot, O., Dehail, P., Aouizerate, B., et al. (2021). Guidelines for the design of a virtual patient for psychiatric interview training. *Journal on Multimodal User Interfaces, 15*(2), 99–107. https://doi.org/10.1007/s12193-020-00338-8

Durak, H. Y. (2023). Conversational agent-based guidance: Examining the effect of chatbot usage frequency and satisfaction on visual design self-efficacy, engagement, satisfaction, and learner autonomy. *Education and Information Technologies, 28*(1), 471–488. https://doi.org/10.1007/s10639-022-11149-7

El Hefny, W., Mansy, Y., Abdallah, M., & Abdennadher, S. (2021). *Jooka: A Bilingual Chatbot for University Admission. In Advances in Intelligent Systems and Computing, (Vol. 1367 AISC,* pp. 671–681). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-72660-7_64

Federici, S., de Filippis, M. L., Mele, M. L., Borsci, S., Bracalenti, M., Gaudino, G., et al. (2020). Inside pandora's box: A systematic review of the assessment of the perceived quality of chatbots for people with disabilities or special needs. *Disability and Rehabilitation: Assistive Technology, 15*(7), 832–837. https://doi.org/10.1080/17483107.2020.1775313

Flowers, J. C. (2018). Strong and weak AI: Deweyan considerations. In *Proceedings of the CEUR workshop proceedings* (p. 2287).

Gonda, D. E., & Chu, B. (2019). Chatbot as a learning resource? Creating conversational bots as a supplement for teaching assistant training course. In *Proceedings of the IEEE international conference on engineering, technology and education (TALE)* (pp. 1–5). https://doi.org/10.1109/TALE48000.2019.9225974

Griol, D., García-Herrero, J., Molina, J. M., & Corchado, J. M. (2011). *The educagent platform: Intelligent conversational agents for E-Learning applications (P. novais, D. preuveneers* (pp. 117–124). Springer Berlin Heidelberg.

Gupta, S., & Chen, Y. (2022). Supporting Inclusive Learning Using Chatbots? A Chatbot-Led Interview Study. *Journal of Information Systems Education, 33*(1), 98–108.

Gutiérrez, Y., Restrepo, E., Baldassarre, M., & Boticar, G. I. O. (2019). Accessibility, biases and ethics in chatbots and intelligent agents for education. In *, 1. Proceedings of the EDULEARN19 proceedings* (pp. 8824–8833). https://doi.org/10.21125/edulearn.2019.2196

Halan, S., Lok, B., Sia, I., & Crary, M. (2014). Virtual agent constructionism: Experiences from health professions students creating virtual conversational agent representations of patients. In *Proceedings of the IEEE 14th international conference on advanced learning technologies, ICALT 2014* (pp. 249–253). https://doi.org/10.1109/ICALT.2014.79. February 2015.

Halan, S., Sia, I., Crary, M., & Lok, B. (2015). Exploring the effects of healthcare students creating virtual patients for empathy training. In *, 9238,. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 239–249). Springer Verlag. https://doi.org/10.1007/978-3-319-21996-7_24.

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

Hayashi, Y. (2013). Pedagogical Conversational Agents for Supporting Collaborative Learning: Effects of Communication Channels. In *Conference on Human Factors in Computing Systems - Proceedings* (pp. 655–660). Association for Computing Machinery. https://doi.org/10.1145/2468356.2468472.

Hobert, S. (2019). Say hello to 'Coding Tutor'! Design and evaluation of a chatbot-based learning system supporting students to learn to program. *In 40th International Conference on Information Systems, ICIS 2019*. Association for Information Systems.

Hien, H. T., Cuong, P.-N., Nam, L. N. H., Nhung, H. L. T. K., & Thang, L. D. (2018). Intelligent Assistants in Higher-Education Environments: The FIT-EBot, a Chatbot for Administrative and Learning Support. In *Proceedings of the 9th International Symposium on Information and Communication Technology* (pp. 69–76). https://doi.org/10.1145/3287921.3287937

Hsu, M. H., Chan, T. M., & Yu, C. S. (2023). Termbot: A Chatbot-Based Crossword Game for Gamified Medical Terminology Learning. *International Journal of Environmental Research and Public Health, 20*(5). https://doi.org/10.3390/ijerph20054185

Hutapea, A. (2017). *CHATBOT : Architecture, design, & development (Tesis. no Usar)*. University of Pennsylvania.

International Organization for Standardisation. (2018). *Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts. in iso 9241-11:2018(E) (Issue november)*. International Organization for Standardisation. https://infostore.saiglobal.com/preview/is/en/2018/i.s.eniso9241-11-2018.pdf?sku=19806670Ahttps://www.sis.se/api/document/preview/80003410/.

Jung, S. (2019). Semantic vector learning for natural language understanding. *Computer Speech & Language, 56*, 130–145. https://doi.org/10.1016/j.csl.2018.12.008

Khalil, M., & Rambech, M. (2022). Eduino: A telegram learning-based platform and chatbot in higher education. *Learning and Collaboration Technologies: Novel Technological Environments, 13329*, 188–204. https://doi.org/10.1007/978-3-031-05675-8_15. LCT 2022, PT II.

Konecki, M., Kadoić, N., & Piltaver, R. (2015). Intelligent assistant for helping students to learn programming. In *Proceedings of the 38th international convention on information and communication technology, electronics and microelectronics, MIPRO 2015 - proceedings* (pp. 924–928). https://doi.org/10.1109/MIPRO.2015.7160406

Kong, J. S. M., Teo, B. S., Lee, Y. J., Pabba, A. B., Lee, E. J. D., & Sng, J. C. G. (2021). Virtual integrated patient: An ai supplementary tool for second-year medical students. *Asia Pacific Scholar, 6*(3), 87–90. https://doi.org/10.29060/TAPS.2021-6-3/SC2394

Krassmann, A., Paz, F., Silveira, C., Tarouco, L., & Bercht, M. (2018). Conversational agents in distance education: Comparing mood states with students' perception. *Creative Education, 09*, 1726–1742. https://doi.org/10.4236/ce.2018.911126

Kumar, H., Yu, K., Chung, A., Shi, J., & Williams, J. J. (2023). Exploring the potential of chatbots to provide mental well-being support for computer science students. SIGCSE 2023. In *, 2. Proceedings of the 54th ACM technical symposium on computer science education* (p. 1339). https://doi.org/10.1145/3545947.3576285.

Kumar, J. A. (2021). Educational chatbots for project-based learning: Investigating learning outcomes for a team-based design course. *International Journal of Educational Technology in Higher Education, 18*(1). https://doi.org/10.1186/s41239-021-00302-w

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., … Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health, 2*(2), Article e0000198. https://doi.org/10.1371/journal.pdig.0000198

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 online tutor in a university course. In *, 2020. Proceedings of the 2020 international symposium on educational technology, ISET* (pp. 53–56). https://doi.org/10.1109/ISET49818.2020.00021

Lee, Y. F., Hwang, G. J., & Chen, P. Y. (2022). Impacts of an AI-based chabot on college students' after-class review, academic performance, self-efficacy, learning attitude, and motivation. *ETR & D-Educational Technology Research and Development, 70*(5), 1843–1865. https://doi.org/10.1007/s11423-022-10142-8

Liu, B. (2021). "Weak AI" is Likely to Never Become" Strong AI", So What is its Greatest Value for us? *arXiv: 2103*, Article 15294.

Lokman, A. S., & Ameedeen, M. A. (2019). Modern chatbot systems: A technical review. *Advances in Intelligent Systems and Computing, 881*, 1012–1023. https://doi.org/10.1007/978-3-030-02683-7_75

Meshram, S., Naik, N., VR, M., More, T., & Kharche, S. (2021). In *Enquiry Chatbot using Rasa Framework, 2021 Asian Conference on Innovation in Technology (ASIANCON)* (pp. 1–8). Pune, India. https://doi.org/10.1109/ASIANCON51346.2021.9544650.

Meyer von Wolff, R., Nörtemann, J., Hobert, S., & Schumann, M. (2020). Chatbots for the Information Acquisition at Universities – A Student's View on the Application Area. In *, 11970. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 231–244). Springer. https://doi.org/10.1007/978-3-030-39540-7_16.

Meyer von Wolff, R., Nörtemann, J., Hobert, S., & Schumann, M. (2020). Chatbots for the Information Acquisition at Universities – A Student's View on the Application Area. In *, 11970. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 231–244). Springer. https://doi.org/10.1007/978-3-030-39540-7_16.

Michos, K., Asensio-Pérez, J. I., Dimitriadis, Y., García-Sastre, S., Villagrá-Sobrino, S., Ortega-Arranz, A., Alario-Hoyos, C., Rodríguez-Triana, M. J., Scheffel, M., Arnedillo-Sánchez, I., Dennerlein, S. M., et al. (2020). *Design of conversational agents for CSCL: Comparing two types of agent intervention strategies in a university classroom BT - Addressing Global challenges and quality education* (pp. 215–229). Springer International Publishing.

Mokmin, N. A. M., & Ibrahim, N. A. (2021). The evaluation of chatbot as a tool for health literacy education among undergraduate students. *Education and Information Technologies, 26*(5), 6033–6049. https://doi.org/10.1007/s10639-021-10542-y

Moldt, J. A., Festl-Wietek, T., Mamlouk, A. M., & Herrmann-Werner, A. (2022). Assessing medical students' perceived stress levels by comparing a chatbot-based approach to the perceived stress questionnaire (PSQ20) in a mixed-methods study. *Digital Health, 8*, Article 20552076221139092. https://doi.org/10.1177/20552076221139092

Moore, N., Ahmadpour, N., Brown, M., Poronnik, P., & Davids, J. (2022). Designing virtual reality-based conversational agents to train clinicians in verbal de-escalation skills: Exploratory usability study. *JMIR Serious Games, 10*(3). https://doi.org/10.2196/38669

Munn, Z., Peters, M., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology, 18*, 143.

Neo, M. (2022). The merlin project: Malaysian students' acceptance of an ai chatbot in their learning process. *Turkish Online Journal of Distance Education, 23*(3), 31–48.

Neo, M., Lee, C. P., Tan, H. Y. J., Neo, T. K., Tan, Y. X., Mahendru, N., et al. (2022). Enhancing students' online learning experiences with artificial intelligence (AI): The MERLIN project. *International Journal of Technology, 13*(5), 1023–1034. https://doi.org/10.14716/ijtech.v13i5.5843

Nimavat, K., & Champaneria, T. (2017). Chatbots: An overview types, architecture, tools and future possibilities. *International Journal of Scientific Research and Development, 5*(7), 1019–1026.

Okonkwo, C. W., & Ade-Ibijola, A. (2021). Python-Bot: A chatbot for teaching python programming. *Engineering Letters, 29*(1), 25–34.

Ondáš, S., Pleva, M., & Hládek, D. (2019). How chatbots can be involved in the education process. In *17th International Conference on Emerging eLearning Technologies and Applications (ICETA), Star ý Smokovec. Slovakia* (pp. 575–580). https://doi.org/10.1109/ICETA48886.2019.9040095

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews, 5*(1), 210. https://doi.org/10.1186/s13643-016-0384-4

Paek, S., & Kim, N. (2021). Analysis of worldwide research trends on the impact of artificial intelligence in education. *Sustainability, 13*(14). https://doi.org/10.3390/su13147941

Perez-Mercado, R., Balderas, A., Munoz, A., Cabrera, J. F., Palomo-Duarte, M., & Dodero, J. M. (2023). ChatbotSQL: Conversational agent to support relational database query language learning. *SoftwareX, 22*. https://doi.org/10.1016/j.softx.2023.101346

Pham, X. L., Pham, T., Nguyen, Q. M., Nguyen, T. H., & Cao, T. T. H. (2018). Chatbot as an intelligent personal assistant for mobile language learning. In *Proceedings of the 2018 2nd international conference on education and e-learning* (pp. 16–21). https://doi.org/10.1145/3291078.3291115

Resch, O., & Yankova, A. (2019). Open knowledge interface: A digital assistant to support students in writing academic assignments. In *Proceedings of the 1st ACM SIGSOFT international workshop on education through advanced software engineering and artificial intelligence* (pp. 13–16). https://doi.org/10.1145/3340435.3342723

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching, 6*(1). https://doi.org/10.37074/jalt.2023.6.1.9

Valdivieso, P. A. C., & Luzoz, M. D. A. (2021). The use of chatbot as an element of tutorial action in university teaching. *REIDOCREA-REVISTA EECTRONICA DE INVESTIGACION Y DOCENCIA CREATIVA, 10*.

Radziwill, N. M., & Benton, M. C. (2017). Evaluating Quality of Chatbots and Intelligent Conversational Agents. *ArXiv*. https://doi.org/10.48550/arXiv.1704.04579

Ryong, K., Lee, D., & Lee, J.-G. (2023). Chatbot's Complementary Motivation Support in Developing Study Plan of E-Learning English Lecture. *International Journal of Human-Computer Interaction*. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85146733413&doi=10.1080%2F10447318.2022.2163786&partnerID=40&md5=7508490e7be787bd3c8d28ad41b9c07f.

Saiz-Manzanares, M., Sánchez, R., Martín Antón, L., González-Díez, I., & Almeida, L. (2023). Perceived satisfaction of university students with the use of chatbots as a tool for self-regulated learning. *Heliyon, 9*, e12843. https://doi.org/10.1016/j.heliyon.2023.e12843

Samyn, K. (2019). 3D Chatbot in Higher Education, Helping Students With Procrastination and Study Planning Problems. In *, 1. EDULEARN19 Proceedings* (pp. 9400–9405). https://doi.org/10.21125/edulearn.2019.2336

Sevgi, U. T., Erol, G., Dogruel, Y., Sonmez, O. F., Tubbs, R. S., & Gungor, A. (2023). The role of an open artificial intelligence platform in modern neurosurgical education: A preliminary study. *Neurosurgical Review, 46*(1). https://doi.org/10.1007/s10143-023-01998-2

Shim, K. J., Menkhoff, T., Teo, L. Y. Q., & Ong, C. S. Q. (2023). Assessing the effectiveness of a chatbot workshop as experiential teaching and learning tool to engage undergraduate students. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-11795-5

Shorey, S., Ang, E., Yap, J., Ng, E. D., Lau, S. T., & Chui, C. K. (2019). A virtual counseling application using artificial intelligence for communication skills training in nursing education: Development study. *Journal of Medical Internet Research, 21*(10), e14658. https://doi.org/10.2196/14658

Sjöström, J., Aghaee, N., Dahlin, M., & Ågerfalk, P. J. (2018). Designing chatbots for higher education practice. In *Proceedings of the SIGED international conference on information systems education and research, October*.

Srilekha, C., Sneha, C., Prathyusha, G. N., Reddy, C. K. K., & Murthy, B. V. R. (2019). Professional study on chatbots using artificial intelligence. *Journal of Applied Science and Computations, VI*(367), 367–378.

Srinivasan, D., Uthayakumar, A., Thiyagalingam, P., Ravindran, S., De Silva, H., & Kasthururarthna, D. (2022). Escort - Natural Language Processing Based University Students Guidance System. In *2022 7th International Conference on Information Technology Research (ICITR)* (pp. 1–6). Sri Lanka: Moratuwa. https://doi.org/10.1109/ICITR57877.2022.9992860.

Srivastava, B. (2021). Did chatbots miss their "Apollo Moment"? Potential, gaps, and lessons from using collaboration assistants during COVID-19. *Patterns, 2*(8), Article 100308. https://doi.org/10.1016/j.patter.2021.100308

Stathakarou, N., Nifakos, S., Karlgren, K., Konstantinidis, S. T., Bamidis, P. D., Pattichis, C. S., et al. (2020). Students' perceptions on chatbots' potential and design characteristics in healthcare education. *Studies in Health Technology and Informatics, 272*, 209–212. https://doi.org/10.3233/SHTI200531

Studente, S., & Ellis, S. (2020). Enhancing the Online Student Experience Through Creating Learning Communities—The Benefits of Chatbots in Higher Education. In S. McKenzie, F. Garivaldis, & K. R. Dyer (Eds.), *Tertiary Online Teaching and Learning*. Singapore: Springer. https://doi.org/10.1007/978-981-15-8928-7_3.

Sweidan, S. Z., Laban, S. S. A., Alnaimat, N. A., & Darabkh, K. A. (2021). SEG-COVID: A student electronic guide within COVID-19 pandemic. In *Proceedings of the 9th international conference on information and education technology (ICIET)* (pp. 139–144). https://doi.org/10.1109/ICIET51873.2021.9419656

Talan, T., & Kalinkara, Y. (2023). The role of artificial intelligence in higher education: ChatGPT assessment for anatomy course. *Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi, 7*(1), 33–40. https://doi.org/10.33461/uybisbbd.1244777

Tan, C. P., Yeap, C. K., Chong, O. L., & Chan, Y. S. (2021). University students' perception on the usefulness of the incorporation of conversational agents in mathematics learning. AICCC 2021. In *Proceedings of the 2021 4TH artificial intelligence and cloud computing conference* (pp. 229–233). https://doi.org/10.1145/3508259.3508292.

Tanana, M. J., Soma, C. S., Srikumar, V., & Atkins, D. (2019). Development and evaluation of ClientBot: Patient-like conversational agent to train basic counseling skills. *Journal of Medical Internet Research, 21*(7), e12529. https://doi.org/10.2196/12529

Wang, Q., Saha, K., Gregori, E., Joyner, D., & Goel, A. (2021). Towards mutual theory of mind in human-ai interaction: How language reflects what students perceive about a virtual teaching assistant. *In Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–14).

Tamrakar, R., & Wani, N. (2021). Design and Development of CHATBOT: A Review. *INTERNATIONAL CONFERENCE On "Latest Trends in Civil, Mechanical and Electrical Engineering*.

Tian, X., Risha, Z., Ahmed, I., Lekshmi Narayanan, A. B., & Biehl, J. (2021). Let's Talk It Out: A Chatbot for Effective Study Habit Behavioral Change. *Proceedings of the ACM on Human-Computer Interaction*. https://doi.org/10.1145/3449171

Tricco, A., Zarin, L. E., O'Brien, K., Colquhoun, H., & Levac, D. (2018). Preferred reporting items for systematic reviews and meta-analyses extension for scoping reviews (PRISMA-ScR) checklist SECTION. *Ann Intern Med, 169*(7), 11–12. https://doi.org/10.7326/M18-0850.2

Turing, A. (1950). Computing machinery and intelligence. *Mind, LIX*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

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

Villegas-Ch, W., García-Ortiz, J., Mullo-Ca, K., Sánchez-Viteri, S., & Roman-Cañizares, M. (2021). Implementation of a virtual assistant for the academic management of a university with the use of artificial intelligence. *Future Internet*, (4), 13. https://doi.org/10.3390/fi13040097

Wambsganss, T., Winkler, R., Söllner, M., & Leimeister, J. M. (2020). A conversational agent to improve response quality in course evaluations. In *Proceedings of the extended abstracts of the 2020 CHI conference on human factors in computing systems* (pp. 1–9). https://doi.org/10.1145/3334480.3382805

Feasibility- Oxford Learner's Dictionaries. feasibility noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com. Oxford University Press. https://www.oxfordlearnersdictionaries.com/definition/english/feasibility.

Yeh, T. P., Chang, S. M., Ho, Y. F., & Ma, W. F. (2021). Online team-based learning teaching strategy for developing caring competencies in nursing students under COVID-19 pandemic restrictions. *Healthcare*, (11), 9. https://doi.org/10.3390/healthcare9111510

Yi, P. K., Ray, N. D., & Segall, N. (2023). A novel use of an artificially intelligent chatbot and a live, synchronous virtual question-and answer session for fellowship recruitment. *BMC Medical Education, 23*(1). https://doi.org/10.1186/s12909-022-03872-z

L. B . T.-FD.D. Zain, S., Baker, D., & Ellis, I. (2021). Digital transformation trends in education. *Chandos digital information review* (pp. 223–234). Chandos Publishing. https://doi.org/10.1016/B978-0-12-822144-0.00036-7

Zaky, Y. A. M. (2023). Chatbot positive design to facilitate referencing skills and improve digital well-being. *International Journal of Interactive Mobile Technologies, 17*(9), 106–126. https://doi.org/10.3991/ijim.v17i09.38395

Zhai, X. (2023). ChatGPT user experience: implications for education. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4312418