Chatbots Evolution in Healthcare: A Systematic Literature Review

Hajar Zerouani*, Abdelhay Haqiq, Bouchaib Bounabat

Abstract—The use of Artificial Intelligence (AI) technology is rapidly expanding across various fields, with chatbots emerging as conversational agents that provide virtual assistance. This paper presents a systematic review that examines the application of chatbots in the healthcare domain. The review aims to explore the evolution of this technology in terms of its demographic and temporal aspects throughout the past years, identify challenges faced, and investigate the factors contributing to successful chatbot implementation in the medical field. To achieve these objectives, a selection of 56 relevant articles from academic databases were analyzed. The findings from this review will provide a comprehensive insight into the utilization of chatbots in healthcare, shedding light on their potential benefits and limitations.

Index Terms—Artificial Intelligence, Chatbot, healthcare, Disease, Systematic Review.

I. INTRODUCTION

Over the past decade, Artificial Intelligence (AI) has transformed the world subtly, machines now can learn and think through without the intervention of humans [1], machines can imitate human cognition, acquiring the ability to learn, think, make decisions, and initiate actions [2]. Chatbot is AI’s extension, it is composed of two words “chat” and “bot” to refer to a bot for messaging that provide a human computer interaction (HCI) [3] which can perform and provide three types of communication’s forms:
- Speech [4] using speech recognition as illustrated in Fig. 1.
- Image based on image processing [5] and machine vision.

According to industry analysts [6], the worldwide chatbot market will be worth 2485.7 USD million by 2028. Nowadays many industries are in the race to develop their chatbots of customer services that offer rapid and smart development [7]. It allows users to receive answers to their questions in a timely manner without having to wait in phone lines or send several emails or search online for responses. With the advancements in artificial intelligence, like healthcare helpdesk systems capable of tackling complex medical issues, chatbots have seen a surge in their prevalence within contemporary society [8].

Back to 1950, Alan Turing’s question ‘can machines think?’ [9] reveals the rise of the chatbot concept. 16 years later, especially in the psychology domain, the first chatbot named ELIZA [10] permits to encourage the patient to talk but the chatbot’s capability were too inflexible, hereafter in 1970 [11] PARRY created to imitate a paranoid patient [12]. Later in 1992 an AI speech synthesis called DR. SBAITSO in psychotherapy [13]. Then ALICE appear in 1995 the name is an acronym of Artificial Linguistic Internet Computer Entity, this chatbot is inspired from ELIZE and designed with pattern matching [14].

The review aims to explore the evolution of this technology in terms of its demographic and temporal aspects throughout the past years, identify challenges faced, and investigate the factors contributing to successful chatbot implementation in the medical field. To achieve these objectives, a selection of 56 relevant articles from academic databases were analyzed. The findings from this review will provide a comprehensive insight into the utilization of chatbots in healthcare, shedding light on their potential benefits and limitations.

The paper is structured as follows: Section 2 discuss the research methodology; Section 3 presents the results obtained from the search process and discusses the results and their implications, and Section 4 concludes the study with suggested future works.

II. REVIEW METHODOLOGY

This research involves examining existing literature on the evolution of CIH and the study follows protocol guidelines for systematic reviews in software engineering outlined...
The guidelines include three phases: plan review, conduct review, and document review, with each phase broken down into several stages as shown in Fig. 2. To address the following structured research questions (SRQs) following the Goal-Question-Metric (GQM) [17]:

SRQ1: Demographics. Are the CIH developed in the most advanced countries technologically or with the best healthcare system?
SRQ2: Application scenarios. What are the different scenarios of CIH used for patients?
SRQ3: Factors. What are the major factors of successful CIH?
SRQ4: Decision making causes. What pushes scientists to work on specific diseases? Is it the most common health condition or the most dangerous one?
SRQ5: Benefits. Which are the advantages provided by the CIHs (for doctors and patients)?
SRQ6: Technology Evolution. Which underlying technologies have been employed to realize CIH?
SRQ7: Drawbacks. Which limitations have CIHs shown?
SRQ8: Open challenges. Which are the open challenges concerning the next generation of CIHs?

Fig. 2. Systematic literature review process (redrawn based on [16]).

A. Plan Review

The primary objective of this research is to uncover pertinent research questions related to the implementation of chatbots in healthcare. The aim is to gain a comprehensive understanding of the most effective practices in developing chatbots for this domain. Additionally, the study delves into the technical attributes of chatbots deployed in the medical sector, providing valuable insights to researchers engaged in chatbot development.

Inclusion criteria: Based on insights from prior research [18], this study will apply specific inclusion & exclusion criteria to select relevant articles.

The following criteria have been identified:
1) Language: Only studies published in the English language will be considered.
2) Publication Date: Articles published before July 2023 will be included in the review.
3) Document Type: Only full-text journal and conference articles will be eligible for inclusion.
4) Focus of Studies: The selection will be limited to studies that center around two listed sets:
   - Contextual keywords: "healthcare," "medical guidance," "clinical assistant," "development," and "evolution".
   - Targeted Keywords: "Conversational agents," "chatbot," and "online assistant.

B. Information source

To gather relevant information for the research topic, the study conducted searches on appropriate articles from digital databases including the following platforms (see Fig 3):
IEEE Digital Library, ScienceDirect, SpringerLink, PubMed, MEDLINE, ACM, EUROPE PMC and IAENG.

Fig. 3. Resources of papers referenced in this systematic review.

C. Conduct Review

The review process involved meticulous research utilizing the specified academic databases and adhering to the established inclusion criteria. Initially, a total of 110 relevant papers were identified and collected. Subsequently, a two-step approach was employed, consisting of coarse-grained and fine-grained examinations. Through this rigorous assessment, 56 primary studies were carefully selected for further analysis, as illustrated in Fig. 4.

D. Document Review

During the final stage of the systematic literature review process, the objective is to examine and retrieve information from chosen articles in order to respond to the SRQs (refer to Fig. 4).

Fig. 4. Conduct Review Process.
III. RESULTS AND DISCUSSIONS

1) SRQ1: Demographics

The analysis of scientific literature from two prominent biomedical databases, Europe PMC and PubMed, reveals a remarkable exponential trend in the interest of the scientific community towards Chatbots in Healthcare (CIH) in the last two years. This trend, as illustrated in Fig. 5, has been particularly evident after the onset of the COVID-19 pandemic. These findings strongly suggest that CIH has the potential to enhance the quality and personalization of medical services, making healthcare more efficient and equitable [19].

Recent statistics from "World Population Review" in 2021 show Asia as the top technologically advanced region, while Europe excels in healthcare systems (Fig. 6). CIH research correlates strongly with countries that possess advanced technology and robust healthcare systems, with Europe and Asia leading in both aspects (Fig. 7 and Fig. 8). Conversely, Africa and South America, despite having excellent healthcare systems, exhibit lower CIH research output due to their technology levels. North America and Oceania, as top technology leaders, demonstrate moderate CIH research activity, influenced by their healthcare system quality. The development of CIH in each country is influenced by both:
- The technological advancement,
- The healthcare system.

2) SRQ2: Application scenarios

As illustrated in Fig. 9, there has been significant scientific interest in mental health, which has resulted in the majority of CIH applications being focused on this area. Mental health encompasses emotional health, psychological health, and social well-being [19]. The following Table I outlines various use cases of CIH in both mental and physical health.

3) SRQ3: Factors

To evaluate the effectiveness and impact of a chatbot, it may be helpful to consider factors such as whether it is supported by healthcare organizations or has partnerships with hospitals. Additionally, gathering statistics on the number of patients who have used the chatbot can provide insights into its reach and impact. Other factors to consider may include the rating of the chatbot on platforms such as the Google Play store and the number of downloads it has received. Considering those major criteria for selecting successful chatbots in Table II:
- Supported from health organizations or it has partner-
ships with hospitals.
- Statistics about how many patients used this chatbot.
- Rating on Google play store and number of downloads. By analyzing these metrics, it may be possible to identify areas for improvement and further optimize the chatbot for the needs of its users.

Here are the common points in the previous CIH:
- **Language relevance**: the use of the English language for chatbots can be an effective way to reach more patients around the world. By having the most widely spoken language available, chatbots can become more accessible to a larger population.
- **Psychological support**: to provide psychological support for individuals with mental health conditions, such as anxiety or depression using sentiment analysis. This involves guiding and assisting individuals by analyzing their emotions and sentiments [42].
- **Collaboration with healthcare organizations**: it is important to collaborate with healthcare organizations, such as hospitals and experts in the healthcare field [40]. This will ensure that the chatbot has access to up-to-date information and guidance from medical professionals, allowing for better communication and interaction with patients. By using AI and regularly updating the chatbot’s data, it can be a valuable resource for individuals in real need of support.
- **Research**: In order to ensure the ongoing development and effectiveness of chatbot software, it is important to conduct research in parallel with the business. This means consistently upgrading and improving the software based on new insights and feedback from users. By doing so, chatbots can continue to provide valuable support and assistance to individuals in need. In order to identify the key elements of successful chatbots, a paper [39] proposes understanding the reasons for chatbot failures in practical applications. The paper lists six common reasons for failure: inadequate resources, subpar content, failure to meet user expectations and poor conversation design, legal regulations, data security and liability concerns, improper use cases, and lack of a business case. Another study [40] shows that the primary technical issues with chatbots are long wait times and excessive demand from multiple users.

### TABLE I
APPLICATION SCENARIO OF CIH

<table>
<thead>
<tr>
<th>References</th>
<th>Application scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>[22-12-24-25]</td>
<td>Self-diagnostic</td>
</tr>
<tr>
<td>[26]</td>
<td>Anamnesis, Medication dosage</td>
</tr>
<tr>
<td>[27][28]</td>
<td>Prioritisation in the emergency room</td>
</tr>
<tr>
<td></td>
<td>Psychiatric treatment, Treatment information</td>
</tr>
<tr>
<td></td>
<td>Food orders for patients</td>
</tr>
<tr>
<td>[28-29-30]</td>
<td>Diet recommendation, Telemedicine</td>
</tr>
<tr>
<td>[31]</td>
<td>Healthcare Education</td>
</tr>
<tr>
<td>[32-33-34]</td>
<td>Medical Consultant</td>
</tr>
<tr>
<td>[35]</td>
<td>Health care Assistant</td>
</tr>
<tr>
<td>[36]</td>
<td>Teleremedy</td>
</tr>
<tr>
<td></td>
<td>Cognitive behavior therapy</td>
</tr>
</tbody>
</table>

4) **SRQ4 and SRQ5**: Decision making causes and Benefits.

According to recent reports from 2019, the top 10 most dangerous diseases in the world are responsible for causing many deaths. The World Health Organization (WHO) classifies ischaemic heart disease [42] as the deadliest disease, followed by trachea, bronchus and lung cancers, Alzheimer’s disease, diarrheal diseases, Diabetes and Kidney diseases. The Fig. 10 shown below compares the number of cases of three diseases mentioned in a recent WHO report, along with COVID-19 which has over 770 million cases worldwide. The study looks at the number of research papers related to chatbots in PubMed and Europe PMC for each disease. This graphical presentation helps us understand how research on chatbots has developed in relation to different diseases.

The Fig. 11 illustrates that there is a greater interest among scientists in cancer research compared to diabetes. This suggests that researchers may not base their decision to work on a specific disease solely on the level of danger posed by the disease. The primary objective of CIH is to enhance efficiency, affordability, convenience, and patient-centered access, with the belief that this will improve health equity and social inclusion [44]. The use of chatbots in the medical field can benefit both individuals and healthcare systems [43]. If chatbots are designed well, they can provide an affordable means for telehealth-based healthcare systems to collect patient data, provide patient education, improve patient engagement, and allocate resources more effectively [45]. In the case of patients suffering from psychological issues, CIH could assist clinicians in relieving pressure, and chatbots could aid doctors in treating these patients by providing support and follow-up for their mental state [46].

**Number of patients in each disease**

![Fig. 10. Number of patients in each disease worldwide by 2023 (see DATA AVAILABILITY).](Image)

**Research in chatbot for 4 diseases**

![Fig. 11. The quantity of studies on chatbots in four specific diseases on December 23, 2023 (see DATA AVAILABILITY).](Image)
5) SRQ6: Technology Evolution

Table III illustrates the progress of CIH over time and how researchers can use each generation to enhance the capabilities of their intelligence programs. The Table III includes several chatbots, beginning with the first chatbot ELIZA [10] in 1966 and ending with the most recent one developed in 2020. By analyzing the evolution of these chatbots, researchers can better understand how technological advancements can benefit their CIH programs.

6) SRQ7: Drawbacks

CIH systems may hinder patient recovery due to confusing or inaccurate responses [21]. While these systems serve well for simple tasks like data collection and administrative work, they lack the capability to handle exceptions and still require supervision by doctors [51]. In cases where a deeper understanding of a patient's condition is necessary, these systems may fail [52]. Furthermore, the absence of rules and regulations safeguarding chatbot users' privacy and security could deter patients from utilizing CIH [52].

The Babylon chatbot has a limitation where doctors are restricted to choosing from diseases included in the chatbot's triage and diagnostic system, potentially leading to biases [53]. These limitations can be categorized into five main issues:

1) Understanding issues: Related to the accuracy of the chatbot's model in comprehending patients' questions or needs effectively.

2) Emotional issues: Concerns regarding the chatbot's ability to respond with appropriate empathy or express sympathy in certain situations.

3) Technical issues: Related to the design of the chatbot's
architecture, robust and safety of machine learning [55] and the technologies utilized.

4) Generating issues: Problems associated with content or data quality, such as dataset size and data update frequency.

5) Privacy issues: Relating to the level of security provided by the chatbot to protect confidential data.

7) SRQ8: Open challenges

In response to the risk of potentially misleading medical treatments arising from flaws in medical diagnostic systems. It is noteworthy that the major challenge for CH is to improve the content of chatbots and test their effectiveness through clinical trials [54]. Despite the significant advancements in AI, chatbots still require the supervision and control of doctors [54]. Since users expect chatbots to communicate and behave like human beings [57], chatbots must learn about human behavior and undergo training in sentimental analysis. Although the most famous chatbot ChatGPT [58] still suffers from a common problem found in many NLP models known as the “hallucination effect”.

Recent research has focused on improving computer-aided diagnosis applications [56] and how to influence users' writing and thoughts using opinionated AI language technologies [59] that will be beneficially especially in health industry. Another research has examined the level of patient engagement with chatbot [60] This paper outlines a path for guiding upcoming research in the creation of patient-centric chatbots tailored for lifestyle and wellness interventions. Despite the potential of chatbots to revolutionize the treatment of chronic illnesses [61] through offering individualized and cost-efficient healthcare, their integration into healthcare systems necessitates a thoughtful evaluation of ethical, legal, and regulatory frameworks.

IV. CONCLUSION

This paper has explored the evolution and challenges of chatbots in the healthcare domain. Over the years, these AI-driven conversational agents have made significant strides, becoming more prevalent in today's healthcare landscape. They have shown immense promise in enhancing patient care, streamlining communication, and providing valuable medical assistance.

Despite efforts to minimize bias and the quality of included studies in this paper, there is still a possibility of subjective judgment in the study selection process, and the selected paper is with free access. This SLR aims to gather, critically appraise, and analyze a wide range of relevant studies to address specific research questions or objectives related to chatbots' applications, benefits, challenges, and impact in healthcare settings. This paper provides insights and recommendations for our future research and development in the field of chatbots in healthcare, as a benchmarking to develop in the future professional chatbot in healthcare.

DATA AVAILABILITY

Figure 5 is based on the data from PubMed and Europe PMC, the collected data is available at https://docs.google.com/spreadsheets/d/13AGK4nLeRWFcXZphP-DAzoX1i0euLjvYu!elzALF0/edit?gid=0.

Figure 6 is based on data from World Population Review, you can see it here: https://worldpopulationreview.com/country-rankings/most-technologically-advanced-countries.

Figure 7 is based on data from The Health Care Index, here to check the data resources: https://ceoworld.biz/2023/08/25_revealed-countries-with-the-best-health-care-systems-2023

Figure 8 is based on data from Scopus, I counted countries of each continents to get result, the collected data is available at https://docs.google.com/spreadsheets/d/13AGK4nLeRWFcXZphP-DAzoX1i0euLjvYu!elzALF0/edit?gid=0.

Figure 9 is based on the data from PubMed, MEDLINE and Europe PMC, the collected data is available at: https://docs.google.com/spreadsheets/d/13AGK4nLeRWFcXZphP-DAzoX1i0euLjvYu!elzALF0/edit?gid=0. I looked for the number of papers interested in the field by 2023. 

Figure 10 is based on the data from PubMed and World Health Organization the collected data is available at https://docs.google.com/spreadsheets/d/13AGK4nLeRWFcXZphP-DAzoX1i0euLjvYu!elzALF0/edit?gid=0.

Figure 11 is based on the data from PubMed and Europe PMC, the collected data is available at: https://docs.google.com/spreadsheets/d/13AGK4nLeRWFcXZphP-DAzoX1i0euLjvYu!elzALF0/edit?gid=0.

REFERENCES


