DOI: 10.36558/rsc.v12i1.7589

Developing chatbots in the field of healthcare: A systematic review

Haniel Gomes Cavalcante
Centro de Ciencias e Tecnologia - CCT
Universidade Estadual do Ceara - UECE
haniel.cavalcante@aluno.uece.br

Thaís de Almeida Barros
Centro de Ciencias e Tecnologia - CCT
Universidade Estadual do Ceara - UECE
thais.almeida@aluno.uece.br

Ricardo Antônio Rebouças Celestino
Centro de Ciencias e Tecnologia - CCT
Universidade Estadual do Ceara - UECE
ricardo.celestino@aluno.uece.br

Daniel Gleison Moreira Lira
Centro de Ciencias e Tecnologia - CCT
Universidade Estadual do Ceara - UECE
daniel.gleison@aluno.uece.br

Francisco Vinicius Nascimento da Silva
Centro de Ciencias e Tecnologia - CCT
Universidade Estadual do Ceara - UECE
francisco.vinicius@aluno.uece.br

Pedro Henrique Araújo de Brito
Centro de Ciencias e Tecnologia - CCT
Universidade Estadual do Ceara - UECE
pedrinho.brito@aluno.uece.br

ABSTRACT

Considering the context of Computer Science, chatbots are computer programs that use Artificial Intelligence techniques to simulate human behavior in dialogues. The use of chatbots applied to the health area has been growing, especially in scenarios for dealing with pandemics, such as COVID-19, as they help to avoid the burden of face-to-face care. Thus, this article proposes a systematic review of the work carried out in this line of research. After the review, it was found which technologies, strategies and frameworks are most used in recent times, as well as which specific areas of health are having more focus on the use of chatbots.

CCS Concepts

· Software and its engineering → Extra-functional properties; Software design engineering; · Applied computing → Health informatics; · Human-centered computing → HCI design and evaluation methods; · Computing methodologies → Natural language processing;

Keywords

chatbots; health; systematic review

1. INTRODUCTION

Artificial Intelligence (AI) is known as a science dedicated to the study of systems that in any observer's perspective, act with intelligence [3]. It can considered a new field of study when compared to other sciences, as it had its beginning in 1956, so some concepts are still being improved. [16] In this context, the chatbot or smart agent are different computing system, where the symbolic and connective approaches can act in a collaborative way, aiming for problem solving. [1]. The basic principle employed in a chatbot consists of an environment that receives questions in human natural language, associates these questions to a knowledge base and finally, emits an answer [5].

Chatbots can be used in many application domains, such as entertainment, business, education and health. Examples of chatbots are the projects ELIZA [21], MGONZ [7], PARRY [6] and ALICE [11]. Regarding chatbot development in the healthcare domain, the works aim to help the interaction between patients and healthcare professionals in many specialities: psychiatry, psychology, pediatrics, cardiology, and many more.

Chatbots used in the healthcare domain can make symptoms mapping and diagnostic predictions, as well as advising instructions for the patient based on machine learning models. By doing so, it is possible to do the care screening remotely and privately, avoiding face-to-face service overload and exposing the patients to unnecessary contact with other patients, especially when dealing with global pandemics, such as COVID-19.

In this context, it is necessary to investigate the use of chatbots in healthcare to determine and identify which techniques, strategies and frameworks are the most used, especially when this use for healthcare is expected to grow in the scientific field, also with empirical evaluations, making it possible to produce systematic reviews of the literature. In this work, research methods, collected results from previous studies and discussions about the development of chatbots in healthcare will be addressed.

2. BACKGROUND

2.1 Technology Applied to Health

Digital health technologies have mobilized governments on the international scene since the early 2000s, [14], resulting in directive documents mobilized by the World Health Organization (WHO), such as the Digital Health Toolbox, launched in 2012 .[14].

In 2018, ahead of the COVID-19 pandemic scenario, governments of several countries unanimously approved a World Health Assembly resolution, calling on the WHO to develop a global digital health strategy to support national efforts to achieve universal coverage. of health. [2].

As a result, the WHO released a document that spells out a global strategy for digital health by the year 2025. [22]. In this document, WHO suggests some characteristics of effective digital technologies in the health scenario. Among them, the following stand out: accessibility, universal access, sustainability, scale in health promotion, disease prevention, diagnosis, management, rehabilitation and palliative care before, during and after epidemics or pandemics, respect for the security and privacy of information about patients' health. [22].

Brazil also has a digital health strategy, with initiatives aimed at telehealth, tele-diagnosis and tele-education.[12]. In line with the WHO guidelines, the Brazilian Government guides the support of other sectors of society, recognizing that the public sector, in isolation, cannot respond to all the needs of scientific discoveries, industrial production, innovation and training of human resources for the Digital Health. [12].

In this strategy, one of the key points is to deal with the difficulties present in emerging countries, such as Brazil. [22]. To increase efficiency in the use of resources in digital health technologies, it is important to invest efforts in research and evaluation of potential global solutions, seeking evidence of their implications for access, cost, quality, safety, sustainability and applicability to different contexts of countries. . [22].

In this scenario, carrying out systematic reviews and mappings can collaborate with the evaluation of studies of eHealth experiences from other countries, sharing successes, failures, trends and good practices. These studies can be used as sources of understanding of the effects that can be achieved and the types of objectives for which eHealth is really relevant, guiding an assertive and efficient allocation of incentive resources. [12].

2.2 Chatbots

2.2.1 Historical Context

Chatbot is a computer program built to simulate the conversations of humans with other humans. It works but receiving an input from the user in natural language, then making a search on its knowledge base for an answer. Finally, it returns the answer trying to imitate human behavior.

This concept came to exist in the 50s, when a British mathematician Alan Turing raised the question: "can machines think?". This problem would be put to test through a game, in a way that it makes it more objective, as described in the article "Computing Machinery and Intelligence" [18].

This game would work the following way: a man (A), a woman (B) and a third person of any sex, where this third person would play the role of interrogator. The interrogator would be put in a different room from the other two people. The objective of the game would be for the interrogator to identify which participant is the male and which is the female. The interrogator would ask questions to both participants, which would receive an answer accordingly.

To avoid influence with the tone of voice, questions would have to be written, without any contact with the participants and the interrogator. The objective of the game is shown when player A is replaced by a machine.

According to Turing, the fundamental question is to know whether the interrogator would have the same success rate after replacing the man for a machine, which would answer the question "can machines think?".

Since the Turing test was proposed, many other advancements took place, such as a program called ELIZA, that was built to win this challenge. This program came to be the base for many more conversational agents that were developed later on.

According to its creator, Joseph Weizenbaum, this program is based mainly in the detection of a keyword that will determine the set of changes applied to the initial sentence [20]. After calculating the answer, it is shown. Therefore, ELIZA doesn't address the question of simulating human intelligence through elaborated techniques, but making use of a simple set of rules and text manipulation which create the illusion of intelligence.

Consequently, a new conversational agent was created, ALICE or Alicebot, which is a variant of ELIZA. Just as ELIZA, ALICE is based on pattern matching mechanisms together with a knowledge base in AIML to elaboratenits answers [19].

The structure of this agent consists of 41,000 elements called categories, where for each category, there is an associated pattern which represents a question and its respective answer. The correspondence between the input text and the defined patterns in AIML are managed by an object that uses few memory, resulting in efficient algorithms that can offer quick responses.

The object responsible for the organization of all the information regarding the categories is called Graphmaster. This information is stored in a tree structure, which allows avoiding duplicated information, obtaining a efficient comprehension of information.

After more technological advancements, in the 2000s, there were other conversational agents created, such as IBM Watson, Microsoft Cortana, Google Assistente, Amaxon Alexa and Samsung S Voice, which represent more developments in the paradigm of chatbots currently. These technologies today are incorporated to various other sophisticated technologies such as voice recognition, natural language understanding and more. [15].

2.2.2 Chatbot Types

Usually, chatbots can be assigned to categories such as in Figure 1 [8]. Those that are retrieval-based work with preset answers and can use languages such as AIML (Artificial Intelligence Markup Language) [19] to manually define interaction patterns previously implemented. Generative-based models, in the other hand, have the ability to generate new answers in real time.

Besides that, another classification that can be utilized for

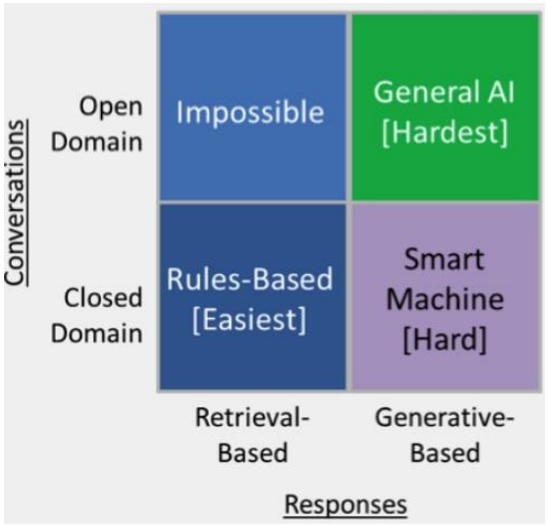

Figure 1: Description of possible chatbot types according to [8].

these systems is about their universe of knowledge in which it proposes to answer questions [13]. Open domain systems aim to answer general questions, based on open data bases from the web. Restricted domain systems address answers to questions of a certain sector, such as the biomedical domain. The closed domain systems are confined to answer questions about a closed collection of documents, which is usually small. Examples of the latter could be a system to answer questions about a Consumer Protection Code or a state's customs laws.

3. METHODOLOGY

For this paper, the purpose of this systematic review is to identify the activities, techniques, methods and tools (frameworks and platforms) considered in the development of chatbots in healthcare. The process used was based on the works of [17] and [9], which provide guidance for systematic reviews in detail. Therefore, the use of this technique has occurred in three phases, as described below.

3.1 Planning

The Planning phase is the once in which the objective of the research must be defined, the way in which the systematic review will be performed, and which criteria will be taken into consideration for the inclusion and exclusion of papers.

3.1.1 Research Objectives

Based on the scenario, as well as in the context that defines the main problem to be addressed in this investigation, the main objective of this research is an analysis of scientific publications that present items of interest related to chatbots in healthcare domain, aiming to determine which technologies, strategies and frameworks are the most used in the area in order to find out the most adequate and safe ones for the healthcare context.

For the systematic review, searches for primary studies published in journal articles and conference proceedings were done, using the electronic search in digital repositories in the areas of healthcare and technology from 2015 to 2021. This interval was chosen so that only recent articles were analyzed, which means there would be higher probability of representing the state-of-the-art.

3.1.2 Research Questions (RQs)

The quality of the systematic review is strongly linked to the proper formulation of the research questions, since the questions guide the review. In this sense, seven questions were elaborated in order to meet the proposed objectives:

- RQ1. What are the tools (*frameworks*/platforms) addressed in the research to provide support?
- RQ2. What is the type of response from *chatbot* according to those described in [8]?
  A. Retrieval-based;
  B. Generative-based;
  C. Other type.
- RQ3. Does the paper propose the use of any technology for the intelligence of the *chatbot* among the most used ones? (single or multiple selection)
  A. The article proposes the use of *Machine Learning*;
  B. The paper proposes the use of *Pattern Matching*;
  C. The paper proposes the use of Ontologies;
  D. The paper does not describe the use of any technology for the development of *chatbots*;
  E. The paper proposes another technology not mentioned in the options above.
- RQ4. What is the adopted/proposed conversational domain according to [8] taxonomy?
  A. The paper proposes the use of Open Domain;
  B. The paper proposes the use of Restricted Domain;
  C. The paper proposes the use of Closed Domain;
  D. The paper does not describe the use of some kind of domain for the development of *chatbots*.
- RQ5. What kind of empirical evaluation was conducted to assess the quality of the *chatbot* ? Which aspect of quality was evaluated?
- RQ6. Which subarea of health care is addressed in the paper?
  A. Psychiatry/Psychology;
  B. Pediatrics;
  C. Nursing;
  D. Physiotherapy;
  E. Another.
- RQ7. What are the challenges and limitations identified in the development of *chatbots*?

3.1.3 Search Strategy for Primary Study Selection

For the analysis and selection of primary studies, the following sources of work were defined: ACM Digital Library, IEEE Xplore Digital Library, ScienceDirect, Elsevier's Scopus and Springer Link, in which the search strings were executed in the title, abstract and keywords containing the following terms: "chatbots", "conversational agent", "conversational bot", "conversational system", "conversational interface", "chat bot", "chatterbot", "chat-bot", "smartbot", "smart bot" , "smart-bot", "virtual coach", "virtual agent", "embodied agent", "relational agent", "avatar", "virtual character", "animated character", "virtual human", "health".

Table 1: Distribution per researcher

Phase	Description	Participating researchers
F1	Execution of the search strategies considering the search strings	5 researchers (one for each base)
F2	Screening: exclusion of primary studies dealing with other issues	2 researchers
F3	First Consensus Meeting	All 7 researchers

The set of search strings for each article base can be found in Appendix A.

3.1.4 Study Selection Procedures

The selective process of this review began with the execution of the search strings. In the second phase, screening took place, where exclusion criteria were applied. At this point, the primary studies were distributed to each researcher according to Table 1. The last phase included a meeting to provide a forum for discussion and consensus among researchers when there were questions for the evaluation of a paper. The purpose of this meeting was to reduce each researcher's bias in order to resolve any doubts in the application of the inclusion/exclusion criteria. In these cases, a complete reading of the doubtful papers was necessary. After this reading, all researchers decided to include or exclude the primary study (PE). The decision was joint to avoid subjectivity.

3.1.5 Criteria for Study Selection

The following inclusion and exclusion criteria were defined to select the primary studies:

a) Inclusion Criteria

The following inclusion criteria for primary studies were defined:

- IC1 - Papers applying chatbots to healthcare;
- IC2 - Papers published from 2015 to 2021;
- IC3 - Papers published in conferences or journals.

b) Exclusion Criteria

The following work exclusion criteria were defined to meet each of the research questions:

- EC1 - Papers in languages other than English;
- EC2 - Papers that have not been published in reputable journals (i.e. journals indexed in the Journal Citation Reports - JCR) or prestigious conferences (i.e. conference level A*, A, B and C categorized in the CORE Conference Ranking);
- EC3 - Papers without full text available;
- EC4 - Papers not related to the development of chatbots in healthcare;
- EC5 - Thesis, books, discussions, opinion papers related to chatbots;
- EC6 - Systematic Reviews.

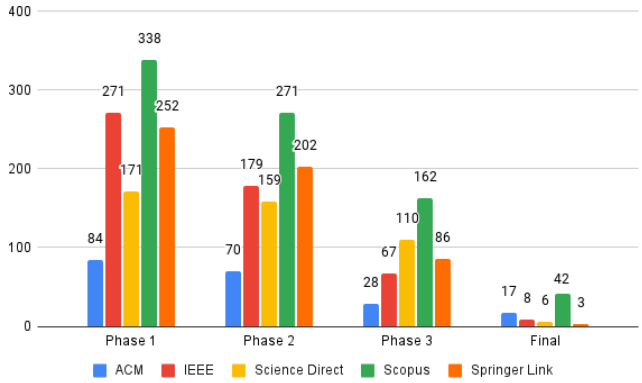

Figure 2: Selection phases chart

3.2 Execution

In phase 1, the automatic search was performed in each digital library. Thus, all documents returned by search queries were included in this phase. In phase 2, articles were considered only in English and with full texts. Papers that were not related to the subject were excluded. This exclusion phase included eliminating duplicate documents, as well as reading the title and abstract. In case of doubt about any article, the paper was included preliminary. The final decision was then considered and evaluated in the next phase. In the third and final phase, the researchers reviewed all articles where there was any uncertainty, performing a complete reading of the article. After the end of this phase, the final list of primary studies that would be analyzed according to the previously defined research questions was defined.

In Figure 2, we can see the works returned at the end of the searches in each phase by search base.

In the first phase, 1116 papers related to the theme of this research were retrieved. Applying the exclusion criteria defined for phase 2, it was possible to identify 881 primary studies. After the consensus meeting of phase 3, 453 relevant articles concerning the objectives of this systematic review remained. After a complete reading of the texts, 76 articles were defined for analysis of the research questions. A flowchart with all the phases can be seen in Figure 3.

The analyzes performed, from the consulted bases, are shown in Figures 4 and 5 and in Table 2.

4. RESULTS ANALYSIS

The final stage of the systematic review was the results analysis. In this stage, each of the 76 articles was analyzed to answer the research questions. In Table 2 we can see the number of articles per selection phase and per search base and in Figure 5, a graph showing the number of these selected files per search base. The list of selected articles, include authors, year of publication and database, is available at Github [1]. In Figure 4 we can see the distribution of selected articles by year. It can be seen that there was an overall tendency of more articles being published on chatbots applied to the healthcare domain, which makes sense as this technology becomes more known and used by people with time, as stated in other works [10]. The sample of pub-

[1]https://github.com/danielgleison/chatbots-health

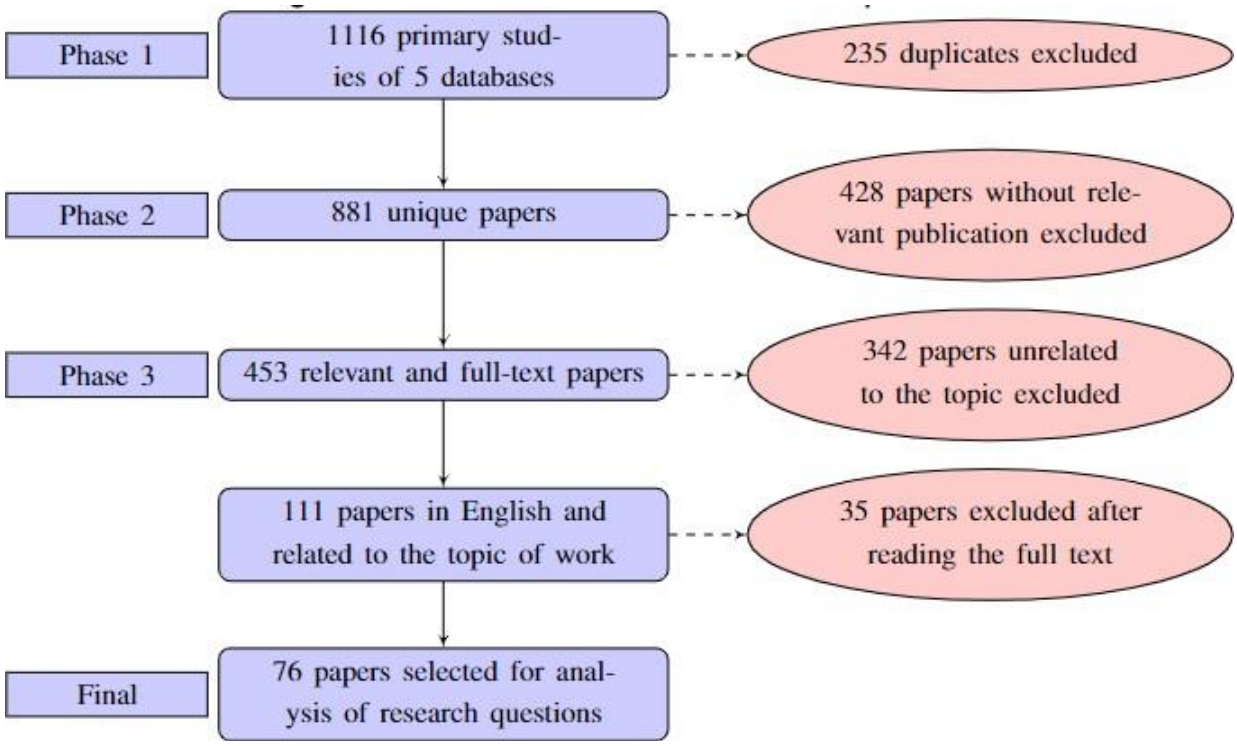

Figure 3: Flowchart of the article selection process

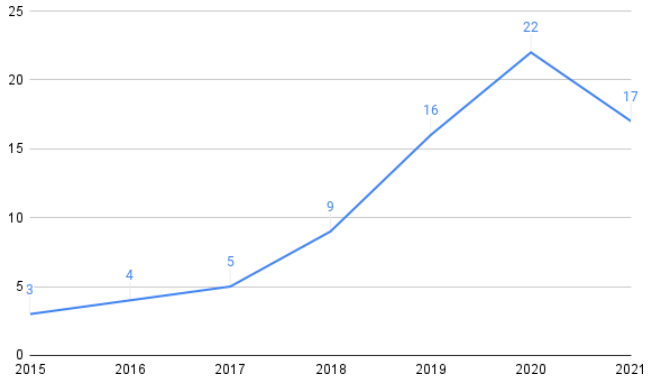

Figure 4: Number of articles per year

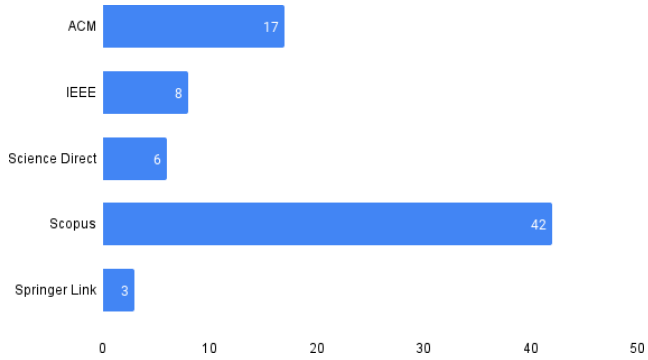

Figure 5: Number of articles per base

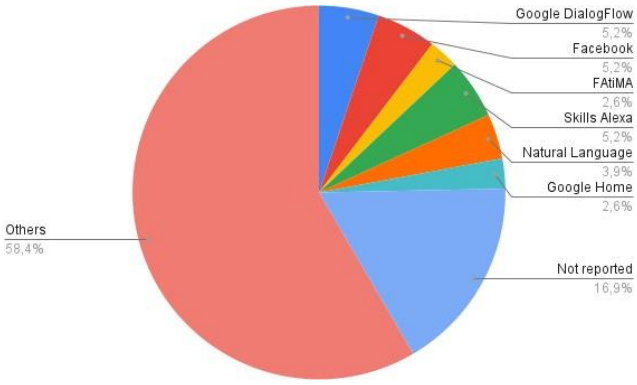

Figure 6: RQ1 response chart

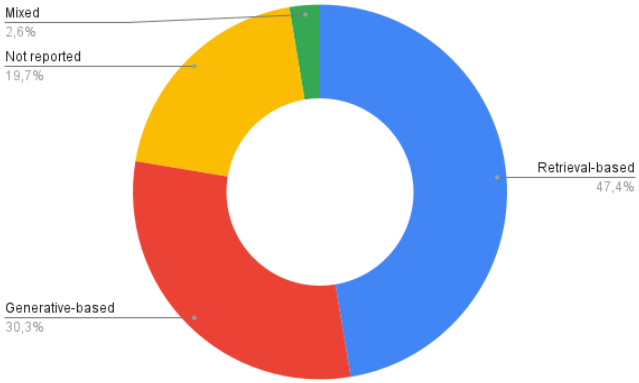

Figure 7: RQ2 response chart

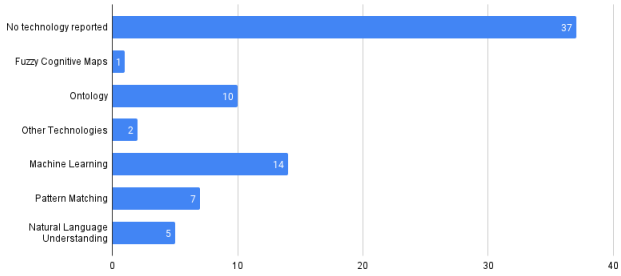

Figure 8: RQ3 response chart

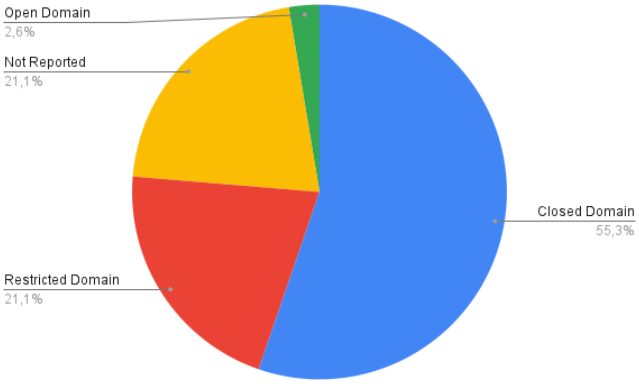

Figure 9: RQ4 response chart

lications was 83% of articles published in conferences and 17% in journals.

Below, we describe the results obtained from each research question.

4.1 Research Question 1

The answers to RQ1. "What are the tools (frameworks / platforms) addressed in the research to provide support?" were as diverse as possible. Here, we can list tools, such as: Google Dialog (5%), Facebook Messenger (5%), Skills from Alexa (5%), FAtiMA (2.6%), among others. In 16.9% of the files reviewed, no tools were reported, as shown in Figure 6.

4.2 Research Question 2

In RQ2. "What is the chatbot response type?", we can see that the chatbot response types are well divided into: retrieval-based (47.4%) and generative-based (30.3%). We have few types of responses that differ from these two ways (2.6%), and in approximately 19.7% of the cases, the authors do not inform the type of response, as shown in Figure 7.

4.3 Research Question 3

In RQ3, "Does the article propose the use of any technology for the intelligence of the chatbot among the most used?", we have a percentage of more than 1/2 of the articles that do not reference any technology. Next, we have the use of Machine Learning (18.4%), the use of ontologies (13.2%) and the use of Pattern Matching (9.2%). Therefore, it is possible to verify the use of the most diverse technologies for the intelligence of the chatbot. This analysis can be verified in Figure 8.

4.4 Research Question 4

For RQ4. "What conversational domain is adopted / proposed?", we can observe that the conversational domains adopted are usually either closed domain (55.3%) or restricted domain (21.1%). In only a little over 2.6% of the files, an open domain is proposed, and in 21.1% of the articles, no domain was proposed or adopted, as shown in Figure 9.

4.5 Research Question 5

To answer RQ5. "What kind of empirical evaluation was performed to evaluate the quality of the chatbot? Which aspect of quality was evaluated?", we observe that there are several forms of empirical evaluation regarding the quality of chatbots. The most common is the analysis based on the feedback provided by the participants with approximately 32.9%. The other most common type of evaluation is based on the collected data, and may or may not use statistical methods, which occurs in approximately 30.3%, and in some cases (11.8%) both types of empirical evaluation are used. Thus, the empirical evaluation is performed in approximately 80% of the articles, whether it is based on the analysis of the collected data, such as the application of statistical methods, K-FOLD cross-validation, BLEU score

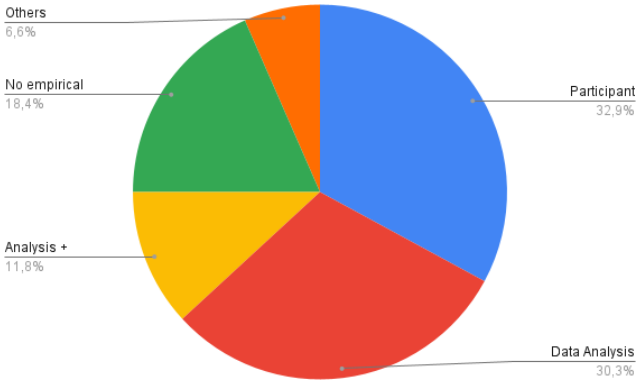

Figure 10: RQ5 response chart

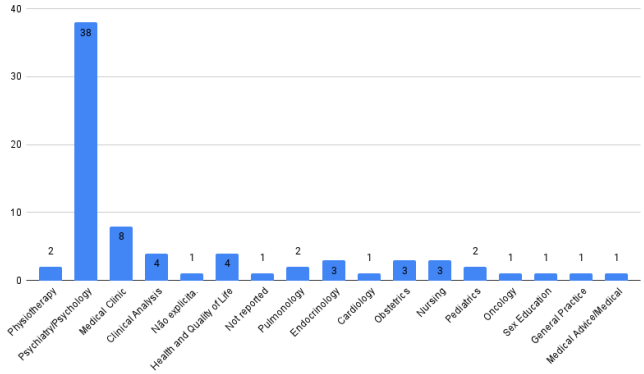

Figure 11: RQ6 response chart

and BERT score, or based on the feedback provided by the participants, where, in most cases, a questionnaire is applied to assess the level of satisfaction, empathy or comfort in the interaction with the chatbot. In addition to these evaluation methods, some articles use both types, achieving a broader evaluation of the tool. In these cases, data such as changes in patient anxiety levels, the degree of intimacy and bonding between the chatbot and the patient, and the user's desire to interact with the chatbot again are evaluated. In 18.4% of the articles, there was no or no report of whether there was any empirical evaluation, according to Figure 10.

4.6 Research Question 6

In RQ6, "Which subarea of the healthcare domain is addressed in the article?" it was observed that the main subarea addressed in the articles is psychiatry / psychology, with approximately half of the occurrences. The subarea of clinical medicine comes in 2nd place with 10.5%, and clinical analysis along with health and quality of life are tied as the third most addressed with 5.3%. In addition, several other areas are addressed, such as obstetrics, pediatrics, endocrinology, and physical therapy. The information is available in Figure 11.

4.7 Research Question 7

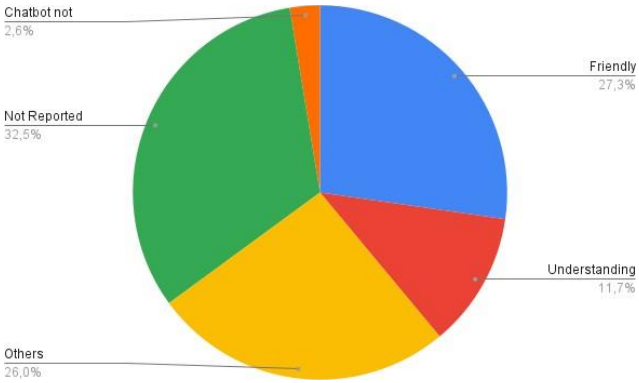

Figure 12: RQ7 response chart

To answer RQ7, "What are the challenges and limitations identified in the development of chatbots?", based on the complete reading of the papers, excerpts were extracted in which the authors identified possibilities for improvement, as well as limitations in the development of chatbots. We detected that, in approximately 32.5% of the analyzed papers, no challenges or limitations regarding the development of chatbots had been reported. In the others, it was verified that the biggest challenge reported would be the need to make the chatbot more friendly and attractive to the user, which was reported in almost 27.3% of the cases. In addition, in 20% of the cases, it was reported that it was necessary to increasingly improve the conversation between the chatbot and the user so that the communication would flow better, according to Figure 12. This makes sense as these challenges are very commonly mentioned in other works [4]. In addition, other challenges/limitations cited were: "Examine the effect of a mental health chatbot on mood in a postpartum population", "Acquire real-life data to improve the algorithm", and "Understand how the use of emotional language influences interaction".

5. THREATS TO VALIDITY

Threats were identified regarding the identification of primary studies, as possible limitations were found in the process of searching for articles that could lead to the absence of related literature and a major challenge of the work was the existence of articles that did not focus on agent engineering per se, but on its interface or only in the empirical evaluation, leaving, in some cases, research questions unanswered. To minimize these threats, leading digital libraries in computing were considered to reduce publication bias. Another threat concerns data extraction, related to possible problems in the data collection phase, such as the subjectivity of the researcher who performs this collection. To reduce this risk, the extraction of information was carried out by a researcher and reviewed by all, in cases of uncertainty.

6. CONCLUSION AND FUTURE WORK

This work performed a systematic review of articles in the healthcare area, where some type of conversational agent is used. Through the proposed research questions, it was possible to determine that the most used chatbot reponse type

is retrieval-based, which is probably because it is usually easier to implement and it works well with a closed collection of documents, which is usually the case. Also, the most used technologies for the intelligence of chatbot are Machine Learning, Ontologies and Pattern Matching that are responsible for around 50%. We can also notice that for chatbots, closed domain and restricted domain are used in more than 80% of the analyzed papers, and it makes sense as the analyzed papers are supposed to have a specific domain (healthcare), and that psychiatry/psychology is the subarea of health that is most addressed in the context, probably because this subarea deals much with conversations and chatbots were designed especially for this task. This shows that a more specific evaluation can be done through more systematic reviews specifically for chatbots in the domain of psychology or psychiatry. This work can be used as a reference for developers looking to implement a conversational agent in the field of healthcare and who want to know the most commonly used technologies, as well as examine different options for strategies and approaches. It is also of interest to researchers, as it maps what is being researched in the area. For future work, other systematic literature reviews can be conducted, this time with a greater focus on a particular sub-area of the field of healthcare or with the use of a particular technology or standard, since this work was quite comprehensive.

7. ADDITIONAL AUTHORS

Additional authors: Mariela Ines Cortes (Centro de Ciencias e Tecnologia - CCT Universidade Estadual do Ceara, email: mariela.cortes@uece.br) .

8. REFERENCES

[1] A. A. Bernardini, A. A. Sônego, and E. Pozzebon. Chatbots: An analysis of the state of art of literature. In *Workshop on Advanced Virtual Environments and Education*, volume 1, pages 1–6, 2018.

[2] O. Brasil. Oms divulga primeira diretriz sobre intervenções de saúde digital, 2019.

[3] B. Coppin. *Inteligência artificial*. Grupo Gen-LTC, 2015.

[4] S. E. Finch, J. D. Finch, D. Huryn, W. Hutsell, X. Huang, H. He, and J. D. Choi. An approach to inference-driven dialogue management within a social chatbot. *arXiv preprint arXiv:2111.00570*, 2021.

[5] L. Fryer and R. Carpenter. Bots as language learning tools. *Language Learning & Technology*, 10(3):8–14, 2006.

[6] J. Huang, M. Zhou, and D. Yang. Extracting chatbot knowledge from online discussion forums. In *IJCAI*, volume 7, pages 423–428, 2007.

[7] M. Humphrys. How my program passed the turing test. In *Parsing the Turing Test*, pages 237–260. Springer, 2009.

[8] Ā. Kamphaug, O.-C. Granmo, M. Goodwin, and V. I. Zadorozhny. Towards open domain chatbots—a gru architecture for data driven conversations. In *International Conference on Internet Science*, pages 213–222. Springer, 2017.

[9] B. Kitchenham and P. Brereton. A systematic review of systematic review process research in software engineering. *Information and software technology*, 55(12):2049–2075, 2013.

[10] L. C. Klopfenstein, S. Delpriori, S. Malatini, and A. Bogliolo. The rise of bots: A survey of conversational interfaces, patterns, and paradigms. In *Proceedings of the 2017 conference on designing interactive systems*, pages 555–565, 2017.

[11] L. A. Lima. Estudo de implementação de um robô de conversação em curso de língua estrangeira em ambiente virtual: um caso de estabilização do sistema adaptativo complexo. 2014.

[12] D. d. I. d. S. Ministério da Saúde, Secretaria-Executiva. Estratégia de saúde digital para o brasil 2020-2028 [recurso eletrônico], 2020.

[13] D. Mollá and J. L. Vicedo. Question answering in restricted domains: An overview. *Computational Linguistics*, 33(1):41–61, 2007.

[14] W. H. Organization. National ehealth strategy toolkit overview. NREL/TP-500-41137, Version 1.21, 2012. Available at https://www.who.int/ehealth/publications/overview.pdf.

[15] M. I. Pavel. Comparing chatbot frameworks: A study of rasa and botkit. 2021.

[16] S. J. Russell and P. Norvig. *Inteligência artificial*. Elsevier, 2004.

[17] N. Sánchez-Gómez, J. Torres-Valderrama, J. A. García-García, J. J. Gutiérrez, and M. J. Escalona. Model-based software design and testing in blockchain smart contracts: A systematic literature review. *IEEE Access*, 8:164556–164569, 2020.

[18] A. Turing and J. Haugeland. *Computing machinery and intelligence*. MIT Press Cambridge, MA, 1950.

[19] R. S. Wallace. The anatomy of alice. In *Parsing the Turing Test*, pages 181–210. Springer, 2009.

[20] J. Weizenbaum. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.

[21] J. Weizenbaum. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 26(1):23–28, 1983.

[22] WHO. Global strategy on digital health 2020–2025, 2020.

APPENDIX

A. SEARCH STRINGS

A.1 ACM Digital Library

```
[[[Publication Title: "chatbot*"] OR
[Publication Title: "conversational agent"] OR
[Publication Title: "conversational bot"] OR
[Publication Title: "conversational system"] OR
[Publication Title: "conversational interface"] OR
[Publication Title: "chat bot*"] OR
[Publication Title: "chatterbot"] OR
[Publication Title: "chat-bot"] OR
[Publication Title: "smartbot"] OR
[Publication Title: "smart bot"] OR
[Publication Title: "smart-bot"] OR
[Publication Title: "virtual coach"] OR
```

```
[Publication Title: "virtual agent"] OR
[Publication Title: "embodied agent"] OR
[Publication Title: "relational agent"] OR
[Publication Title: "avatar"] OR
[Publication Title: "virtual character"] OR
[Publication Title: "animated character"] OR
[Publication Title: "virtual human"]] AND
[Publication Title: "health*"]] OR [[[Keywords: "chatbot*"]
OR [Keywords: "conversational agent"]
OR [Keywords: "conversational bot"]
OR [Keywords: "conversational system"]
OR [Keywords: "conversational interface"]
OR [Keywords: "chat bot*"]
OR [Keywords: "chatterbot"]
OR [Keywords: "chat-bot"]
OR [Keywords: "smartbot"]
OR [Keywords: "smart bot"]
OR [Keywords: "smart-bot"]
OR [Keywords: "virtual coach"]
OR [Keywords: "virtual agent"]
OR [Keywords: "embodied agent"]
OR [Keywords: "relational agent"]
OR [Keywords: "avatar"]
OR [Keywords: "virtual character"]
OR [Keywords: "animated character"]
OR [Keywords: "virtual human"]]
AND [Keywords: "health*"]]
OR [[[Abstract: "chatbot*"]
OR [Abstract: "conversational agent"]
OR [Abstract: "conversational bot"]
OR [Abstract: "conversational system"]
OR [Abstract: "conversational interface"]
OR [Abstract: "chat bot*"]
OR [Abstract: "chatterbot"]
OR [Abstract: "chat-bot"]
OR [Abstract: "smartbot"]
OR [Abstract: "smart bot"]
OR [Abstract: "smart-bot"]
OR [Abstract: "virtual coach"]
OR [Abstract: "virtual agent"]
OR [Abstract: "embodied agent"]
OR [Abstract: "relational agent"]
OR [Abstract: "avatar"]
OR [Abstract: "virtual character"]
OR [Abstract: "animated character"]
OR [Abstract: "virtual human"]]
AND [Abstract: "health*"]]
AND [Publication Date: (01/01/2015 TO 12/31/2021)]
```

A.2 IEEE Xplore

```
((((("Document Title": "chatbots"
OR "Document Title": "conversational agent"
OR "Document Title": "conversational bot"
OR "Document Title": "conversational system"
OR "Document Title": "conversational interface"
OR "Document Title": "chat bot"
OR "Document Title": "chatterbot"
OR "Document Title": "chat-bot"
OR "Document Title": "smartbot"
OR "Document Title": "smart bot"
OR "Document Title": "smart-bot"
OR "Document Title": "virtual coach"
OR "Document Title": "virtual agent"
OR "Document Title": "embodied agent"
OR "Document Title": "relational agent"
OR "Document Title": "avatar"
OR "Document Title": "virtual character"
OR "Document Title": "animated character"
OR "Document Title": "virtual human")
OR ("Abstract": "chatbots"
OR "Abstract": "conversational agent"
OR "Abstract": "conversational bot"
OR "Abstract": "conversational system"
OR "Abstract": "conversational interface"
OR "Abstract": "chat bot"
OR "Abstract": "chatterbot"
OR "Abstract": "chat-bot"
OR "Abstract": "smartbot"
OR "Abstract": "smart bot"
OR "Abstract": "smart-bot"
OR "Abstract": "virtual coach"
OR "Abstract": "virtual agent"
OR "Abstract": "embodied agent"
OR "Abstract": "relational agent"
OR "Abstract": "avatar"
OR "Abstract": "virtual character"
OR "Abstract": "animated character"
OR "Abstract": "virtual human")
OR ("Author Keywords": "chatbots"
OR "Author Keywords": "conversational agent"
OR "Author Keywords": "conversational bot"
OR "Author Keywords": "conversational system"
OR "Author Keywords": "conversational interface"
OR "Author Keywords": "chat bot"
OR "Author Keywords": "chatterbot"
OR "Author Keywords": "chat-bot"
OR "Author Keywords": "smartbot"
OR "Author Keywords": "smart bot"
OR "Author Keywords": "smart-bot"
OR "Author Keywords": "virtual coach"
OR "Author Keywords": "virtual agent"
OR "Author Keywords": "embodied agent"
OR "Author Keywords": "relational agent"
OR "Author Keywords": "avatar"
OR "Author Keywords": "virtual character"
OR "Author Keywords": "animated character"
OR "Author Keywords": "virtual human") )
AND ("Document Title": "health"
OR "Abstract": "health"
OR "Author Keywords":"health") NOT ("systematic review")))
```

A.3 Science Direct

```
("chatbot" OR "conversational" OR "smartbot"
OR "coach" OR "agent" OR  "avatar"
OR "character" OR "virtual human")
AND ("health") Title, abstract, keywords. Year: 2015-
2021
```

A.4 Elsevier's Scopus

```
(TITLE-ABS-KEY ( "conversational agent"
OR "conversational bot"
OR "conversational system"
OR "conversational interface"
OR "chatbot"
OR "chat bot"
```

```
OR  "chat-bot"
OR "chatter bot"
OR "chatterbot"
OR "smartbot"
OR "smart bot" OR "smart-bot" )
OR TITLE-ABS-KEY ( "virtual agent"
OR "virtual character"
OR "virtual coach"
OR "virtual human"
OR "Avatar"
OR "Embodied agent"
OR "relational agent"
OR "Animated character" ) )
AND TITLE-ABS-KEY ("Health" )
AND DOCTYPE ( cp )
AND PUBYEAR > 2014
AND ( LIMIT-TO ( SUBJAREA,"COMP" ) )
AND ( LIMIT-TO ( LANGUAGE,"English" ) )
AND ( LIMIT-TO ( SRCTYPE,"p" )
OR LIMIT-TO ( SRCTYPE,"j" ) )
```

A.5 Springer Link

```
"conversational agent"
OR "conversational bot"
OR "conversational system"
OR "conversational interface"
OR "chatbot" OR "chat bot"
OR "chat-bot"
OR "chatter bot" OR "chatterbot"
OR "smart bot" OR "smartbot"
OR "smart-bot" OR "virtual agent"
OR "virtual character"
OR "virtual coach" OR "virtual human"
OR "avatar" OR "embodied agent"
OR "relational agent" OR
"animated character" AND "health"
```