

# Usability of Chatbots: A Systematic Mapping Study

Ranci Ren

Dep. Ing. Informática  
Univ. Autónoma de Madrid  
Madrid, Spain  
ranci.ren@estudiante.uam.es

John W. Castro

Dep. Ing. Informática y Ciencias de la Computación  
Universidad de Atacama  
Copiapó, Chile  
john.castro@uda.cl

Silvia T. Acuña, Juan de Lara

Dep. Ing. Informática  
Univ. Autónoma de Madrid  
Madrid, Spain  
{silvia.acunna, juan.delara}@uam.es

**Abstract— Background:** The use of chatbots has increased considerably in recent years. These are used in different areas and by a wide variety of users. Due to this fact, it is essential to incorporate usability in their development. **Aim:** Our objective is to identify the state of the art in chatbot usability and applied human-computer interaction techniques, to analyze how to evaluate chatbots usability. **Method:** A systematic mapping study has been conducted, searching the main scientific databases. The search retrieved 170 citations and 19 articles were retained as primary studies. **Results:** The works were categorized according to four criteria: usability techniques, usability characteristics, research methods and type of chatbots. **Conclusions:** Chatbot usability is a very incipient field of research, where the published studies are mainly surveys, usability tests, and rather informal experimental studies. Hence, it becomes necessary to perform more formal experiments to measure user experience, and exploit these results to provide usability-aware design guidelines.

**Keywords—**Usability; chatbots; systematic mapping study

## I. INTRODUCTION

Chatbots are computer programs with a textual or voice interface, based on natural language [1]. They are specifically designed to make user interaction as natural as possible, and have received extensive attention from academia and industry in recent years. Chatbots not only enable a faster and more natural way to access information, but they will become a key factor in the process of humanizing machines in the near future.

Usability is defined as the degree to which a program can be used to achieve quantified objectives with effectiveness, efficiency, and satisfaction in a specified context of use [2]. Usability is a critical aspect in interactive software systems and so it is essential to incorporate usability in chatbots, to improve user experience. Chatbots are become pervasive and are used in many areas, such as bookings of all sorts of services, to obtain medical advice and for online shopping [1][3][4]. The multiple uses and benefits of chatbots explain their strong growth in terms of users, satisfaction and saving resources. It is expected that the number of users will grow in the US by 23.1% [5]. Although the market is still beginning to take shape (compared to the number of websites, the number of bots is still not large) it is estimated that the market size will expand massively [4].

Many universities and commercial companies have put into use chatbots interacting with mature systems. At the commercial level, Facebook messenger already has more than 300,000 chatbots in use [5]. This makes downloading and installing new apps unnecessary, and the use of smartphones allows for personalization possibilities [6]. Further, the use of

chatbots can be more cost-effective than human-assisted support [7]. Some companies are building chatbots independently (e.g., Microsoft is promoting the idea of “*conversation as a Platform*”) to support a variety of media, from Skype to search [8]. Chatbots are not an emerging concept. Research on dialogue systems can be traced back to the 50s, where Alan M. Turing posed the question “*can machines think?*” proposing the Turing test as a criterion for judging whether the machine has intelligence [9]. Weizenbaum’s development of ELIZA at MIT in the 60s can be considered the first dialogue system [1]. Lately, the advances in natural language processing (NLP) have boosted the raise of many chatbot development frameworks (e.g. DialogFlow (<http://dialogflow.com>)).

However, there are currently few works that discuss the usability of chatbots in an integrated and formalized manner. The objectives of our research are to identify the state of the art in chatbots usability and the applied Human-Computer Interaction (HCI) techniques by a Systematic Mapping Study (SMS) and to analyze how to evaluate the usability of chatbots. The contribution of this research is a picture of the current state of usability in chatbots. For this purpose, we present a SMS where we classify the types of chatbots, the measured usability characteristics, the applied usability techniques and the research methods used to evaluate chatbot usability.

**Paper organization.** In Sec. 2, we present related work. In Sec. 3, we describe the research method of the SMS. Sec. 4 presents the results of the SMS. In Sec. 5, we discuss the results and threats to validity, and finally Sec. 6 concludes.

## II. RELATED WORKS

We found only three systematic reviews related to chatbots [9][10][11]. The one by Klopfenstein et al. [9] surveys conversational interfaces, patterns, and paradigms. However, they do not detail the literature retrieval process, and hence it may be potentially incomplete. The survey traces the history of chatbots, from ELIZA to modern chatbots for MOOCs. They conclude that only a subset of chatbots are designed for communicating in natural language, which sometimes makes users disappointed. Then they compare features of major messaging platforms that support bots, like Messenger, WeChat, Line and Skype. Most of them already support a variety of message types, pictures, videos, and sounds. However, none of them have comprehensive enough features. For example Line has groups, buttons and carousel features, but no payment and quick message reply. They detail advantages of bots for users and developers, and conclude

stressing the benefits of chatbots as a new software platform to provide services and data to users.

The work by Ramesh et al. [10] surveys design techniques for conversational agents. The paper presents various solutions for building chatbots, including AIML, NLP and Natural Language Understanding (NLU). The authors describe the general structure of a chatbot, which consists of a responder, classifier and graph master. Then, they list several design techniques for chatbots, from pattern matching, to recurrent neural networks. They stress that NLP techniques are increasingly being used in recent years. The paper presents a classification of chatbots, which includes retrieval-based and generative-based, long and short conversations and open/closed domain. Retrieval-based chatbots pick responses from a pool of predefined ones. The third work, by Laranjo et al. [11] makes a systematic review of conversational agents in health. This review retrieved 1,513 research papers, and identified 17 primary studies. The search was performed in April 2017 and updated in February 2018. They describe 14 different conversational agents distinguishing type of communication technology, dialogue management, dialogue initiative, input modality and task-oriented aspects. The evaluation measures were divided into three main types: technical performance, user experience and health-related measures.

Overall, these works do not focus on usability techniques or usability characteristics of chatbots. Therefore, to the best of our knowledge, there is no SMS on the status of the chatbot usability. Our work covers this gap.

### III. RESEARCH METHOD

We aim to answer the following research questions: **(RQ1)** *What is the state of the art of usability in the development of chatbots?* and **(RQ2)** *How to evaluate the usability of chatbots using HCI principles?* To answer both questions, we have executed an SMS to identify and classify these issues in the published literature [12].

#### Search String Selection

The first step is identifying search strings and relevant keywords. For this purpose, several options were tried and the best one chosen. In particular, we first read some initial articles, obtaining keywords and basic knowledge related to the topic. After combining the opinions of two experts in HCI, we opted for the search string: (usability OR “usability technique” OR “usability practice” OR “user interaction” OR “user experience”) AND (chatbots OR “chatbots development” OR “conversational agents” OR chatterbot OR “artificial conversational entity” OR “mobile chatbots”).

#### Databases and Search Protocol

The search was performed in sequence from Scopus, ACM Digital Library, IEEE Xplorer, SpringerLink and Science Direct. The search fields used were determined by the options provided by each database. Considering that the concept of chatbots is still relatively new, the search range is from January 2014 to October 2018. We ordered the search considering the data base that returned most results. The search fields were selected to assure that searches were similar across data bases.

The criteria used to retrieve the fundamental studies are summarized below.

- **Inclusion criteria:** The paper is written in English; AND The abstract or title mentions an issue regarding the chatbots and usability; OR The abstract mentions an issue related to usability engineering or HCI techniques; OR The abstract mentions an issue related to the user experience.
- **Exclusion criteria:** The paper does not present any issue related to the chatbots and usability; OR The paper does not present any issue related to the chatbots and user interaction; OR The paper does not present any issue related to the chatbots and user experience.

#### Paper Selection

The searches were run using the search string defined. The number of papers returned by the first search was 170, which are called *Retrieved Papers*. Then by inspecting the title, keywords and abstract of each retrieved paper, 41 papers were filtered to the group of *Candidate Papers*. The whole group of Candidate Papers was screened for duplicates. When duplicates were found, only the first occurrence of the paper was counted and maintained, the others were deleted. The final group has 39 papers, which is called *Non-Duplicate Candidate Papers*. Each paper of the Non-Duplicate Candidate Papers group was read, to determine if they described any sort of usability of chatbots. The results were cross-checked by two experts in the HCI area, and any disagreement was discussed and resolved in our meetings. Finally, 19 papers were identified as primary studies.

### IV. SYNTHESIS OF THE RESULTS

Figure 1 provides an overview of the primary studies retrieved by the SMS. It is made of three categories, determined by the year of publication, type of paper (conference, journal, chapter) and usability characteristics. The left-hand side is composed of two scatter (XY) charts with bubbles at the intersections of each category. The size of each bubble is determined by the number of primary studies that have been classified as belonging to the respective categories at the bubble coordinates.

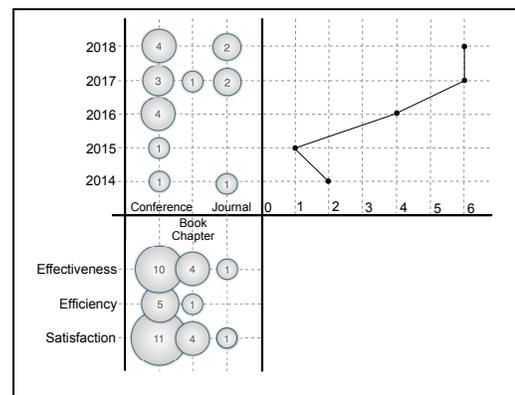


Figure 1. Overview of the Primary Studies

The right-hand side of the Figure shows the number of primary studies by publication year. Publications started to grow from 2015, and many articles (mainly in conferences)

have been published each year since then, confirming the interest in the field. It can be noted that most interest in chatbot usability is on *effectiveness* and *satisfaction*.

After conducting the SMS and analyzing the literature with respect to the usability of chatbots, the primary studies were classified from four different perspectives: usability techniques, usability characteristics, research methods and type of chatbots. These categories are reviewed next.

### A. Usability Techniques

The primary studies in this category identify the adoption of usability techniques from HCI. This is the second-most studied group in the literature. The usability techniques are shown in Table I. From the analysis of the papers, we found that *questionnaires* and *interviews* are most commonly used.

TABLE I. USABILITY TECHNIQUES

Usability Techniques	Primary studies
Questionnaire (SUS/ad-hoc)	[1][3][13][14][15][16][17][18][19][20][21][22][23][24][25][26]
Interview	[15][21][22][23][24][26][27][28][29]
Think-aloud	[15][19][22]
Direct observation	[20]
Cognitive walkthrough	[22]

In most cases, two or more techniques are combined for the usability evaluation. Each of these methods has its own characteristics, and cannot fully meet all requirements of the usability test in isolation. Hence, it is necessary to combine various methods. For example, in [20], direct observation and the *System Usability Scale* (SUS) questionnaire are jointly used. In [21], questionnaires and interviews are used together in every research phase. In [22] the authors conduct a usability test to compare the usability of three chatbot platforms by using a SUS questionnaire, think-aloud and interview to rate the feedback from participants. A post-task questionnaire and an open-ended interview were used together in [26]. In [15], they video-recorded the experiment process for a retrospective think-aloud, and then conducted an interview and a questionnaire after accomplishing the tasks. In [24], user testing is combined with questionnaires and interview. In [23], they conducted semi-structured interviews and used different standard questionnaires together in the first and last evaluation period. In the middle period, to gather more comprehensive information, they used the SUS questionnaire. In [13], they developed a pre-study questionnaire to illustrate the types of interactions perceived to be the most frequent with an Alexa chatbot. In [1], participants were asked to fill the metrics to measure their user experience, and also were asked to compare two different interfaces and justify their responses.

In some cases, the authors used just one single technique to measure usability. In [14], they conducted a survey using a questionnaire. In [16], though they mainly used questionnaires to measure usability, they track user experience through different questionnaires from different periods with open questions. In [27], they used structured interviews. In [17], different questionnaires were used in three different periods of

the experiment. In [18], the authors used questionnaires to measure quantitative and qualitative evaluations of the new NLP method used by the chatbot. In [29], during the interview, participants had to explain the difficulties they had. In [28] to avoid excessive verbosity and to use verbal instead of text feedback, they used interview with open questions. In [25], to evaluate their web client, the authors used a questionnaire related to reliability, usability, and functionality of the system. Overall, we can conclude that the technique used depends on the specific conditions, while there is no standard proposal.

### B. Usability Characteristics

According to the primary studies, usability characteristics are mainly identified in three aspects: Effectiveness, Efficiency and Satisfaction.

1) *Effectiveness*: Effectiveness is defined as the accuracy and completeness with which users achieve specified goals in HCI [30][31]. From Figure 1, most papers consider effectiveness as an essential factor when evaluating the usability of chatbots. Table II shows more details on the used effectiveness criteria.

TABLE II. EFFECTIVENESS

Measures of Effectiveness	Primary studies
Task completion	[1][13][22][25][27]
Accuracy	[17][18][24][25][26][28]
Recall	[18][25]
Experts and Users' assessment	[14][15][21][24]

We have identified task completion, accuracy of chatbot reply, comparison with recall and expert assessment as the main means to assess effectiveness. In [14] by gathering feedback from experts and potential users, they evaluate grading of the perceived quality of effectiveness of the chatbot [31] and find some shortcomings and possible solutions that will enhance the application's usability for its intended audience. In these works, the number of correct responses or interventions indicates the accuracy (to measure if user achieve specified goal [30]) and recall (users' ability to recall information from the interface [31]). The result shows that most chatbots achieve the required accuracy and recall of response [25][26]. For example, through comparing with other chatbots with similar functionality for completing the task, the authors in [1] proved their e-commerce chatbot performs better than the default chatbot. In [18], according to the result of the questionnaire, 80% participants claimed that the content of the retrieved information is clear and useful. In [27], the authors measure the number of users who complete the task (interview) through two different tools, showing that the chatbot has higher acceptability. To identify the measures of characteristics accuracy and recall, the works [30][31] have been followed.

However, there are still problems to achieve a high level of task completion and accuracy of the chatbot reply. In [13], 19 incomplete tasks were reported, because of an ill-defined system design. In [17], during the evaluation, there were some problems with the DBpedia semantic entry point, which

affected the accuracy of some of the users. In [28], 46 entries were negotiated, of which 7 (15.2 %) did not correspond correctly to the user’s original wishes, but when a participant use more lengthy sentences, he produced noticeably more utterances compared to the average of the others. This problem mainly resulted from the inability of the system to process long, convoluted utterances properly and lacking the ability to guide the user during the interaction. In [24], the chatbot generated unnecessary information in response to highly structured conversations. In [25], the factor affecting the accuracy of chatbot reply is the need to handle one or more user conversation turns before providing the answer.

2) *Efficiency*: Efficiency relates to the resources expended in relation to the accuracy and completeness with which the users achieve their goals [30][31]. Most papers discuss *task completion time*, *mental effort* and *communication effort* to use the chatbot, as shown in Table III.

TABLE III. EFFICIENCY

Measures of Efficiency	Primary studies
Task completion time	[1][29]
Mental effort	[1][3]
Communication effort	[17][20][21]

In [1], the authors compare the number of views and average time the participants took in completing a task with the Convey chatbot, and a default one. The results showed they spent more effort and time with Convey in performing the task.

Perceived autonomy and competence are factors favoring efficiency in chatbot usability [3]. In [17], it was noted that, since the chatbot can correct erroneous inputs, users do not need to spend much communication effort when talking to the chatbot. In addition, less communication effort makes the chatbot easier to operate. In [20] the authors count the number of participants’ cumulative assertions to measure the communication effort. Its steady increase demonstrates that users can use the chatbot efficiently in short time. Finally, users spend more communication effort when the chatbot has limited conversational ability, as discussed in [21].

3) *Satisfaction*: This is the largest group of papers within the primary studies. Satisfaction is defined as the degree to which user needs are satisfied when a product or system is used in a specified context of use [30][31]. The measures of satisfaction include *ease-of-use*, *context-dependent questions*, *satisfaction before* and *during use*, *complexity control*, *physical discomfort* of the interface, *pleasure*, the *willing of use* the chatbot *again*, and enjoyment and *learnability*. From Table IV, the ease-of-use, willing to use the chatbot again and user experience are the main measures of satisfaction used. Emotional aspects such as perceived utility, pleasure, comfort, are also considered in [24], and are related to the user experience. Among the primary studies, works have been found measuring the user experience mainly considering the physical discomfort and pleasure. These works are highlighted with a rectangle in the Table IV.

It must be noted that chatbots have more exploration space for interaction with users. A physical chatbot was proposed in

[16] to support self-management of diabetes by children. The usability evaluation included capabilities, social presence, and the quantity of speech and movements. Children stated that physical chatbots were more (inter)active, more present and capable of doing different things, such as dancing. Chatbots with actual images or entities are more likely to establish relationships with users, improving their experience. In [29], a combination of speech-and-gesture makes users get well better with the chatbot. In [22], the authors compared Pandorobot with two other chatbots. Overall, Pandorobot’s voice sounded less robotic, entertaining users better.

TABLE IV. SATISFACTION

Measures of Satisfaction	Primary studies
Ease-of-use	[1][14][16][17][18][19][22][23][24][26]
Context-dependent question	[22][23][29]
Before use	[14][21][23]
During use	[3][21]
Complexity control	[14][18][23][25]
Physical discomfort	[14][18][21]
Pleasure	[1][13][16][17][21][22][26][29]
Want to use again	[1][14][16][17][21]
Learnability	[15][16][22]

Besides, more flexibility and speech commands context-dependent are required for better usability. In [1] participants mentioned that a shopping chatbot was easy to use since it tracked their search history. In [17] some users did not consider they needed an affective enhanced semantic chatbot at home. In [23] the authors observed that the acceptance of the chatbot decreases since its response mismatched the users’ initial expectations. Potential explanations for such inconsistencies might include fundamental differences in user expectations for the chatbot and the emphasis on the interactive and entertaining qualities of the system over its informational value. The user background should be considered a key point in evaluating satisfaction. Cultural, socio-economical and personal preferences can influence the opinions towards the chatbot. In [23], the authors noticed that users in the Netherlands were more experienced with technology than in the other two countries of the study, therefore their expectations towards the novel technology were higher. In [15] users with higher technical knowledge learned quicker to use the chatbot.

### C. Research Methods

The research methods used by the authors of the primary studies within this group include surveys of chatbots users’ experience, experiments of using chatbots to realize some given tasks, usability tests, case studies and quasi-experiments. The research methods are detailed in Table V. The most common research methods include *survey*, *experiment* and *usability test*. In most experiments, very simple tasks are proposed. For example, using Apple Siri to find an inexpensive hotel in Osaka [15], search a flight ticket and hotel [3], whether

a simple chatbot can be appropriately used as a delivery mechanism [24], buying shoes [1], measuring the quality and quantity of the information retrieved [18], taking a structured interview with chatbot [27], or playing a game [29]. However, real-life situations are more complicated.

TABLE V. RESEARCH METHODS

Research methods	Primary studies
Survey	[13][16][19][22][24][25][26]
Experiment	[1][3][13][14][15][16][17][19]
Usability test	[18][19][22][25][27][28]
Case study	[14][20][23][29]
Quasi-experiment	[20][21][25]

Rather than aiming to fully recreate the real-world task, simulation-based assessment should incorporate psychologically relevant aspects and situations from the real-world task and the environment, such as time-pressure, or high uncertainty. In [17], the authors show that when the chatbot has visual appearance and emotions, users do not notice small changes in voice and facial expressions. The experiment concludes that it is not necessary to use extremely accurate facial expressions for realistic use. In [21], the authors deployed a digital pet avatar in the participants’ home for 3 months to simulate real-life situations as much as possible. In [20], experiments were designed in a complex way, to simulate real-world situations. This is sometimes necessary, because if we were unable to use the chatbot effectively with a design as realistic as possible (but nonetheless simplified), it would be unlikely to be effective under more challenging conditions for the military, law enforcement and others in safety-critical real-world environments.

In most cases, researchers make comparisons. For example, in [14], the authors compare a chatbot with a similar one or with a similar application. Research methods can also be combined, which typically yields better results. In [13], a pre-survey questionnaire was performed to assess the usability of an Alexa chatbot. Then, an experiment was conducted to investigate specific problems. Machine learning (ML) algorithms, in combination with cloud-based databases can be used to solve some current shortcomings of handling natural language (e.g., chatbots can’t recognize the words that they haven’t been programmed for, and some chatbots speak unnatural language [24][26]).

#### D. Type of Chatbots

The AIML technology is still widely used in the design of chatbots [24][25]. However, the use of chatbots using NLP is growing [20]. For example, the PAL project [16] can generate reasonable feedback through user-entered information. In [13], the authors show that the chatbot can be used via natural language phonic control, to perform search, entertainment, and to control other devices. In [14], more users are satisfied with the chatbot, due to its speaker functionality and natural conversation flow. In [18], the authors used Object Relational Mapping (ORM) frameworks to improve the process of

generating SQL statements from NL queries. Many chatbots are built as Embodied Conversational Agents (ECA), and there are increasing number of chatbots with image, sound and personality [17][21][23][29]. However, sometimes chatbots have negative emotions. When the ECA has a negative personality, it tends to ignore or blame the user [17]. In addition, the chatbot is required to have the ability to learn and adapt to its user context to be useful [23]. Therefore, complete evaluations should be carried out to obtain a better comprehension of these issues.

## V. DISCUSSION AND VALIDITY THREATS

The analysis reveals that the incorporation of usability techniques in the chatbot development process in a formalized manner is not strongly reflected in the primary studies. We found three papers reviewing the chatbot literature: one discussing the conversational interfaces, patterns, and paradigms [9], one investigating design techniques for conversational agents [10], and a systematic review of conversational agents in healthcare [11]. None of them does a SMS in chatbots usability, which proves our work is original. Judging by the increase in publications since 2015, the integration of usability of chatbots is of notable interest. However, there is no agreement on what would be a formalized and more systematic integration yet. Therefore, it is an open problem that requires more research effort. Even though the literature retrieved by the SMS provides a picture of chatbot usability, no paper provides generally applicable guidelines for chatbots usability. On one hand, the validity of our SMS is threatened by including only papers written in English. On the other, the authors of an SMS may make errors of judgement when analyzing the relevant publications. This is a horizontal rather than a vertical analysis, on which ground relevant publications may have been overlooked. Additionally, although the terms used in the search string were the most commonly accepted by other authors, other terms used describing relevant publications may have been overlooked. Finally, the publications were evaluated and classified based on the judgment and experience of the authors, and other researchers may have evaluated the publications differently.

## VI. CONCLUSIONS

This paper has described an SMS study conducted to answer two research questions: **RQ1.** *What is the state of the art of usability in the development of chatbots?* We retrieved 19 primary studies dealing with integration from four different perspectives: usability techniques, usability characteristics, research methods and types of chatbots. The usability techniques are applied to evaluate the usability of the developed chatbot, but not in the analysis and design activities of the chatbot. The procedure more frequently followed to evaluate the usability of chatbot is to select a group of subjects to use the chatbot freely or perform certain tasks and then measure satisfaction with a SUS survey.

**RQ2.** *How to evaluate the usability of chatbots using HCI principles?* The evaluation of the usability of chatbots must be done considering the context of use, i.e. the environment where the chatbot will be used, and with representative subjects to whom the chatbot is directed. The most commonly used

methods are surveys, experiments and usability tests. The experimentation and replication of experiments is key within HCI. Achieving successful replicas in a discipline allows its results to be added to previous ones, making the body of knowledge grow. However, there is an absence of controlled experiments and replicas measuring chatbots usability.

There are many ways for practitioners to apply the usability material in this paper: (i) The chatbot implementation team can use usability characteristics (Tables II-IV) as checklists to help them solve critical problems, and (ii) comparing the test results of the same system at different times can check whether the usability characteristics is improved or decreased.

The real-life application of a chatbot will save time to companies, leading to financial gain because of the tasks it is able to take on. As the intelligence and technology of chatbots evolve, they will take on more responsibilities. The chatbot industry is very much interested in the adoption of usability techniques in its development process. On this ground, there is a need for usability-aware design guidelines.

#### ACKNOWLEDGMENT

Work supported by the R&D program of the Madrid Region (S2018/TCS-4314), and the Spanish Ministry of Science (project MASSIVE, RTI2018-095222-B-I00).

#### REFERENCES

- [1] M. Jain, R. Kota, P. Kumar and S.N. Patel. "Convey: Exploring the use of a context view for chatbots". Proc. CHI Conference on Human Factors in Computing Systems, p. 468. 2018.
- [2] ISO 9241-11. "Ergonomic requirements for office work with visual display terminals (VDTs)—Part II guidance on usability". 1998.
- [3] Q.N. Nguyen and A. Sidorova. "Understanding user interactions with a chatbot: A self-determination theory approach". In Americas Conference on Information Systems 2018: Digital Disruption. 2018.
- [4] K.Panetta. "Gartner's top 10 strategic technology trends for 2017". Online: <https://www.gartner.com/smarterwithgartner/gartner-top-10-technology-trends-2017/>. [Accessed march 26-2019]
- [5] J. Pereira and O. Diaz. "A quality analysis of facebook messenger's most popular chatbots". Proc. of ACM SAC, pp. 2144-2150. 2018.
- [6] C. Messina. "2016 Will be the year of conversational commerce". Online: <https://medium.com/chris-messina/2016-will-be-the-year-of-conversational-commerce-1586e85e3991>. 2016. [Accessed 12/14/2018].
- [7] J. Lester, K. Branting and B. Mott. "Conversational agents". The Practical Handbook of Internet Computing, Chapman and Hall/CRC, pp. 220-240. 2004.
- [8] L. Sullivan. "Facebook chatbots hit 70% failure rate as consumers warm up to the tech". Online: <https://www.mediapost.com/publications/article/295718/> 2017. [Accessed march 26-2019]
- [9] L.C. Klopfenstein, S. Delpriori, S. Malatini and A. Bogliolo. "The rise of bots: A survey of conversational interfaces, patterns, and paradigms". Proc. Conference on Designing Interactive Systems, pp. 555-565. 2017.
- [10] K. Ramesh, S. Ravishankaran, A. Joshi and K. Chandrasekaran. "May. A survey of design techniques for conversational agents". In Int. Conf. on Inform., Commun. and Computing Technology, pp. 336-350. 2017.
- [11] L. Laranjo, A.G. Dunn, H.L.Tong, A.B.Kocaballi, J.Chen, R.Bashir, D.Surian, B.Gallego, F.Magrabi, A.Y. Lau and E.Coiera. "Conversational agents in healthcare: a systematic review". Journal of the American Medical Informatics Associat., 25(9), pp.1248-1258. 2018.
- [12] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson. "Systematic mapping studies in software engineering". In Proceedings of the 12th Int. Conf. on Evaluation and Assessment in Soft. Eng., pp. 71-80. 2008.
- [13] I. Lopatovska, K. Rink, I. Knight, K. Raines, K. Cosenza, H. Williams, P. Sorsche, D. Hirsch, Q. Li, and A. Martinez. "Talk to me: Exploring user interactions with the amazon alexa". Journal of Librarianship and Information Science. 2018.
- [14] A. Cheng, V. Raghavaraju, J. Kanugo, Y.P. Handrianto and Y. Shang. "Development and evaluation of a healthy coping voice interface application using the Google home for elderly patients with type 2 diabetes". In Consumer Communications & Networking Conference (CCNC), 2018 15th IEEE Annual, pp. 1-5. 2018.
- [15] M.L. Chen and H.C. Wang. "How personal experience and technical knowledge affect using conversational agents". Proc. 23rd Inter. Conf. on Intelligent User Interfaces Companion, p. 53. 2018.
- [16] C. Sinoo, S. van der Pal, O.A.B. Henkemans, A. Keizer, B.P. Bierman, R. Looije and M.A. Neerinx. "Friendship with a robot: Children's perception of similarity between a robot's physical and virtual embodiment that supports diabetes self-management". Patient education and counseling. 2018.
- [17] J. Pérez, Y. Sánchez, F.J. Serón, and E. Cerezo. "Interacting with a semantic affective ECA". Proc. Conf. Intelligent Virtual Agents, pp. 374-384. 2017.
- [18] A. Alghamdi, M. Owda and K. Crockett. "Natural language interface to relational database (NLI-RDB) through ORM". Advances in Intelligent Systems and Computing, vol 513. Springer, Cham 2017.
- [19] M.L. Tielman, M.A. Neerinx, R. Bidarra, B. Kybartas and W.P. Brinkman. "A therapy system for post-traumatic stress disorder using a virtual agent and virtual storytelling to reconstruct traumatic memories". Journal of medical systems, 41(8), p.125. 2017.
- [20] A. Preece, W. Webberley, D. Braines, E.G. Zaroukian and J.Z. Bakdash. "SHERLOCK: Experimental evaluation of a conversational agent for mobile information tasks". IEEE Transactions on Human-Machine Systems, 47(6), pp.1017-1028. 2017.
- [21] N.C. Chi, O. Sparks, S.Y. Lin, A. Lazar, H.J. Thompson and G. Demiris. "Pilot testing a digital pet avatar for older adults". Geriatric Nursing, 38(6), pp.542-547. 2017.
- [22] J. Saenz, W. Burgess, E. Gustitis, A. Mena and F. Sasangohar. "The usability analysis of chatbot technologies for internal personnel communications". In 67th Annual Conference and Expo of the Institute of Industrial Engineers, pp. 1357-1362. 2017.
- [23] C. Tsiouri, J. Quintas, M. Ben-Moussa, S. Hanke, N.A. Nijdam and D. Konstantas. "The CaMeLi Framework—A multimodal virtual companion for older adults". In Proceedings of SAI Intelligent Systems Conference, pp. 196-217. 2018.
- [24] D. Elmasri and A. Maeder. "A conversational agent for an online mental health intervention". Proc. Int. Conf. Brain and Health Informatics, pp. 243-251. 2016.
- [25] A.I. Niculescu, K.H. Yeo, L.F. D'Haro, S. Kim, R. Jiang and R.E. Banchs. "Design and evaluation of a conversational agent for the touristic domain". In APSIPA, pp. 1-10. 2014.
- [26] S. Tegos, S. Demetriadis and T. Tsiatsos. "A configurable conversational agent to trigger students' productive dialogue: a pilot study in the CALL domain". International Journal of Artificial Intelligence in Education, 24(1), pp.62-91. 2014.
- [27] J.A. Micoulaud, P. Sagaspe, E. De Sevin, S. Bioulac, A. Sauteraud and P. Philip. "Acceptability of embodied conversational agent in a health care context". Proc. Conf. Intelligent Virtual Agents, pp. 416-419. 2016.
- [28] R. Yaghoubzadeh, K. Pitsch and S. Kopp. "Adaptive grounding and dialogue management for autonomous conversational assistants for elderly users". Proc. Conf. Intelligent Virtual Agents, pp. 28-38. 2015.
- [29] D. Novick and L.M. Rodríguez. "Extending empirical analysis of usability and playability to multimodal computer games". In Intern. Conf. of Design, User Experience, and Usability, pp. 469-478. 2016.
- [30] ISO/IEC 25010. "Systems and software engineering—systems and software quality requirements and evaluation (SQuaRE)—system and software quality models". 2010.
- [31] K. Hornbæk. "Current practice in measuring usability: Challenges to usability studies and research". International Journal of Human-Computer Studies, 64(2), pp. 79-102. 2006.