Article:

Abd-Alrazaq, AA, Alajlani, M, Abdallah Alalwan, A et al. (3 more authors) (2019) An overview of the features of chatbots in mental health: A scoping review. International Journal of Medical Informatics, 132. 103978. ISSN 1386-5056

https://doi.org/10.1016/j.ijmedinf.2019.103978

An overview of the features of chatbots in mental health: A scoping review

Alaa A Abd-alrazaq[a]; Mohannad Alajlani[b]; Ali Abdallah Alalwan[c]; Bridgette M Bewick,[d]; Peter Gardner[e]; Mowafa Househ[a*]

[a] College of Science and Engineering, Hamad Bin Khalifa University, Qatar

[b] Institute of Digital Health, University of Warwick, United Kingdom

[c] Amman University College for Banking and Financial Sciences, Al-Balqa Applied University, Jordan

[d] Leeds Institute of Health Sciences, School of Medicine, University of Leeds, United Kingdom

[e] School of Psychology, University of Leeds, United Kingdom

* Corresponding author.

E-mail address: mhouseh@hbku.edu.qa

Abstract

Background: Chatbots are systems that are able to converse and interact with human users using spoken, written, and visual languages. Chatbots have the potential to be useful tools for individuals with mental disorders, especially those who are reluctant to seek mental health advice due to stigmatization. While numerous studies have been conducted about using chatbots for mental health, there is a need to systematically bring this evidence together in order to inform mental health providers and potential users about the main features of chatbots and their potential uses, and to inform future research about the main gaps of the previous literature.

Objective: We aimed to provide an overview of the features of chatbots used by individuals for their mental health as reported in the empirical literature.

Methods: Seven bibliographic databases (Medline, Embase, PsycINFO, Cochrane Central Register of Controlled Trials, IEEE Xplore, ACM Digital Library, and Google Scholar) were used in our search. In addition, backward and forward reference list checking of the included studies and relevant reviews was conducted. Study selection and data extraction were carried out by two reviewers independently. Extracted data were synthesised using a narrative approach. Chatbots were classified according to their purposes, platforms, response generation, dialogue initiative, input and output modalities, embodiment, and targeted disorders.

Results: Of 1039 citations retrieved, 53 unique studies were included in this review. Those studies assessed 41 different chatbots. Common uses of chatbots were: therapy (n=17), training (n=12), and screening (n=10). Chatbots in most studies were rule-based (n=49) and implemented in stand-alone software (n=37). In 46 studies, chatbots controlled and led the conversations. While the most frequently used input modality was writing language only (n=26), the most frequently used output modality was a combination of written, spoken and visual languages (n=28). In the majority of studies, chatbots included virtual representations (n=44). The most common focus of chatbots was depression (n=16) or autism (n=10).

Conclusion: Research regarding chatbots in mental health is nascent. There are numerous chatbots that are used for various mental disorders and purposes. Healthcare providers should compare chatbots found in this review to help guide potential users to the most appropriate chatbot to support their mental health needs. More reviews are needed to summarise the evidence regarding the effectiveness and acceptability of chatbots in mental health.

Keywords

Chatbots, Conversational agents, Mental health, Mental disorders, Depression.

Abbreviations

AA: Alaa Abd-Alrazaq

MA: Mohannad Alajlani

MH: Mowafa Househ

YLDs: Years Lived with Disability

1 Introduction

The World Health Organisation defines mental health as “a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community” [1]. It is has been estimated that mental disorders may affect 29% of people in their lifetime [2]. Globally, mental disorders are considered one of the most common causes of disability [3]. In 2010, 28.5% of global Years Lived with Disability (YLDs) were caused by mental, neurological and substance use disorders, making them the top cause of YLDs [4]. Further, they caused about 10% of global Disability-Adjusted Life Years (DALYs) [4]. Importantly, absolute DALYs for these disorders increased from 182 million to 258 million (41%) over 20 years (1990-2010) [4]. It has been estimated that the global economy will lose $16 trillion between 2011 and 2030 through lost labour and capital output resulted from mental disorders [5].

There is a global shortage of mental health workers, with demand out-stripping service provision [6]. Specifically, while developed countries have only about 9 psychiatrists per 100,000 people [7], low income countries have as few as 0.1 for every 1,000,000 people [8]. Due to the relative lack of mental health resources, it is difficult to provide mental health interventions using the one-on-one traditional gold standard approach [5]. According to the World Health Organization, mental health services do not reach about 55% and 85% of people in developed and developing countries, respectively [9]. The lack of access to mental health services may lead to suicidal behaviour, resulting in increasing mortality [10].

The insufficient number of mental health workers has prompted the utilization of technological advancement to meet the needs of the people who are affected by mental health conditions [5]. According to the World Health Organization [9], 29% of the 15,000 mobile health apps focused on mental health. One of the main technological solutions to the lack of mental health workforce are chatbots, also known as conversational agents, conversational bots, and chatterbots [6].

A chatbot is a system that is able to converse and interact with human users using spoken, written, and visual languages [6, 11]. Chatbots have potential to increase access to mental health interventions. In particular, chatbots may encourage interaction by those who have traditionally been reluctant to seek mental health advice due to stigmatization [6, 12]. Many chatbots have been developed for providing mental health interventions. For example, the chatbot “Wysa” uses several evidence-based therapies (e.g. cognitive behavioural therapy, behavioural reinforcement, and mindfulness) to target symptoms of depression for users [13]. LISSA is another chatbot that provides training for people with autism in order to develop their social skills [14].

Numerous studies have been conducted to assess different aspects of using chatbots for mental health, such as effectiveness [15, 16], acceptability [14, 17], usability [18, 19], and adoption [20, 21]. Bringing this evidence together is very important to inform mental health providers and users about the main features of chatbots and their potential uses, and to inform future research about the main gaps of the previous literature. Two reviews have been conducted. The first, a scoping review focused on only embodied conversational agents (i.e. chatbots that show virtual human characters on their screens to mimic main features of human face-to-face conversation, such as verbal and nonverbal behavior) [22]. The second review focused on both embodied and non-embodied conversational agents (i.e. chatbots that communicate with users via only texts appears on screens and do not show virtual human characters), but it focused on some mental disorders (i.e. depression, anxiety, schizophrenia, bipolar, and substance abuse disorders) [6]. That review used limited search terms and therefore identified a relatively low number of studies (i.e. 10) [6]. Thus, it is necessary to conduct a review that focuses on all types of chatbots (embodied and non-embodied) for mental health using accurate and comprehensive search terms. Accordingly, the aim of the current review was to provide an overview of the features of chatbots used by individuals for their mental health as reported in the empirical literature.

2 Methods

To achieve the abovementioned objective, a scoping review was carried out. A scoping review is defined as an initial exploration of the available research literature which follows a systematic method in order to map evidence on a certain area and identify its scope, size, and nature [23]. This scoping review followed guidelines recommended by the PRISMA Extension for Scoping Reviews (PRISMA-ScR) [23].

2.1 Search strategy

2.1.1 Search sources

For the purpose of the current study, we searched the following bibliographic databases: MEDLINE, EMBASE, PsycINFO, Scopus, Cochrane Central Register of Controlled Trials, IEEE Xplore, ACM Digital Library, and Google Scholar. Only the first 100 citations resulted from searching Google Scholar were scanned in this review. This is because Google Scholar usually retrieves several hundred citations which are ordered by their relevance to the search topic. The search was conducted from the 10th to the 12th of May 2019. Reference lists of the included studies and reviews were checked for additional studies of relevance to the review (backward reference list checking). Further, we checked relevant studies that cited the included studies using the “cited by” function available in Google Scholar (forward reference list checking).

2.1.2 Search terms

Search terms considered in the current review were selected based on two elements: population (e.g. mental health, mental disorder, mood disorder, and anxiety disorder) and intervention (e.g. conversational agent, chatbot, chatterbot, and virtual agent). The search terms were derived from previous reviews and informatics experts interested in mental health. Further, search terms for mental disorders were derived from the Medical Subject Headings (MeSH) index in MEDLINE. Appendix A shows the search strings used for searching each electronic database.

2.2 Study eligibility criteria

In order for studies to be included, they had to convey primary research findings regarding chatbots used by individuals for their mental health. The review focused on chatbots that work on the following platforms: stand-alone software and web browser, but not robotics, serious games, SMS, nor telephones. We excluded studies containing chatbots that were designed to be used specifically by physicians or caregivers. Studies about chatbots whose dialogue was generated by a human operator were excluded. The review included peer-reviewed articles, dissertations, conference proceedings, and reports, but not reviews, conference abstracts, proposals, editorials. Studies had to be written in the English language to be included in the review. There were no restrictions regarding the type of dialogue initiative (i.e. use, system, mixed), input and output modality (i.e. spoken, visual, and written), study design, study setting, measured outcome, year of publication, and country of publication.

2.3 Study selection

The current review followed two steps in selecting the studies. In the first step, two reviewers (AA & MA) screened independently the titles and abstracts of all retrieved studies. In the second step, the same reviewers read independently the full texts of studies included from the first step. Any disagreements between both reviewers were resolved through consulting a third reviewer (MH). Inter-coder agreement between both reviewers was assessed using Cohen's kappa [24], which was 0.80 and 0.84 in the first and second step of the selection process, respectively, indicating a very good agreement [25].

2.4 Data extraction

To conduct a systematic and accurate extraction of data, we developed a data extraction form and piloted it using five included studies (Appendix B). Similar to the study selection process, two reviewers (AA & MA) independently conducted the process of data extraction, and any disagreements were resolved by the third reviewer (MH). Inter-coder agreement between the reviewers was good (Cohen's kappa = 0.74) [25].

2.5 Study quality assessment

It is well known that scoping reviews are different from systematic reviews in having broader topics and including studies with more diverse study designs [26, 27]. Therefore, scoping reviews usually do not focus on the quality assessment of the included studies [26, 27]. Accordingly, we did not assess the quality of the included studies in this review.

2.6 Data Synthesis

Extracted data were synthesised using a narrative approach. We endeavoured to classify the chatbots according to their purpose, platforms, response generation, dialogue initiative, input and output modalities, embodiment, and targeted disorders (Appendix B). To that end, we adapted several taxonomies published in the literature [28-30]. Characteristics of studies and population were summarised in a table and described narratively. Then, a description of the characteristics of chatbots in the included studies was presented.

3 Results

3.1 Search results

As shown in Figure 1, the search process of six bibliographic databases retrieved 1039 citations. After removing 409 duplicates, 630 unique titles and abstracts remained. After scanning their titles and abstracts, 505 citations were excluded. By reading the full text of the 125 remaining citations, 43 publications were included. Six additional studies were identified from backward reference list checking, and four studies were identified from forward reference list checking. In total, 53 publications were included in the synthesis. Appendix C shows the full list of the included studies.

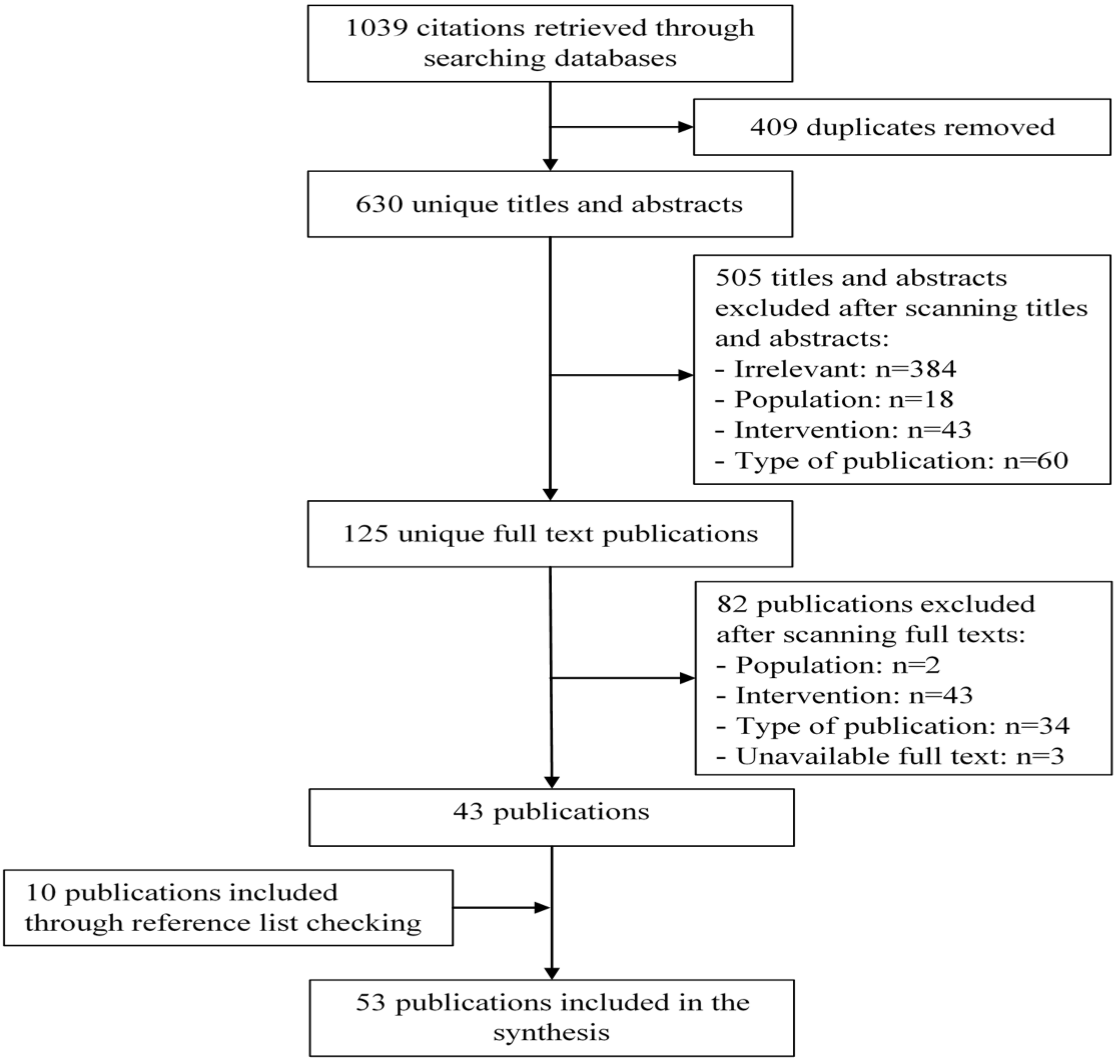

Figure 1: Flow chart of the study selection process

3.2 Description of included studies

As presented in Table 1, the most commonly used study design was quasi-experimental (n=23, 43%). About two thirds of studies were journal articles (n=35, 66%). Although studies were published in more than 17 countries, about 43% of studies were published in the USA (n=23). More than half of the studies were published between 2016 and 2019 (n=30, 56.6%). The sample size was less than 50 in 30 studies whereas it was 200 or more in only 4 studies. The mean sample size was 75.2, it ranged from 4 to 454. Age of participants was reported in 33 studies, the mean age of participants in all those studies was 37.1 years. Sex of participants was reported in 41 studies, the mean of male percentage in all those studies was 55%. A clinical sample relating to a specific mental disorder featured in about 58.5% of the studies (n=31). Participants were recruited from either clinical settings (n=20), community (n=16), or educational settings (n=20). The most common outcomes assessed by the included studies were acceptability (n=36) and effectiveness of the chatbots (n=33). Appendix D shows a description of each included study.

Table 1: Characteristics of the included studies.

Characteristics	**Number of studies**
Study design	Quasi-experiment:23 Survey:16 Randomised controlled trial:14
Type of publication	Journal article:35 Conference proceeding:16 Thesis:2
Country	USA:23 Japan:6 Australia:4 Netherlands:3 France:3 UK:3 Turkey:1 Germany:1 Sweden:1 Greece:1 Spain:1 Korea:1 Pakistan:1 Global population:1 China: 1 Spain & Mexico:1 Romania, Spain and Scotland:1
Year of publication	≤2005:0 2006-2010:7 2011-2015:15 2016-2019:31
Sample size	<50:30 50-99:10 100-199:9 ≥200:4
Mean age	37.1[1] years
Sex (male)	55%[2]
Sample type	Clinical sample:31 Non-clinical sample:22
Setting[3]	Clinical:21 Educational:20 Community:16 Unknown:3
Measured outcomes[4]	Acceptability:36 Effectiveness:33 Usability:20 Adoption:7
Tips	[1]: Mean Age was reported in 33 studies.

	[2]: Sex was reported in 41 studies. [3]: Numbers do not add up as some studies recruited the sample from more than one setting. [4]: Numbers do not add up as most studies assessed more than one outcome.
Abbreviations	UK: United Kingdom; USA: United State of America

3.3 Description of chatbots

The included studies assessed 41 different chatbots. The chatbots were given names in 39 studies (e.g. Woebot, Wysa, Laura). In 17 studies, chatbots were used for therapeutic purposes (Table 2). For example, the chatbots “Woebot” and “Help4Mood” were developed to deliver cognitive behavioural therapy for patients with depression and anxiety [31, 32]. In other 12 studies, chatbots were used for training purposes. For instance, the chatbots “LISSA” and “VR-JIT” were used to train patients with autism to improve their social skills and job interview skills, respectively [14, 33]. Chatbots were used in ten studies as a screening tool for several disorders such as dementia [34-36], tobacco and alcohol use disorders [37], stress [15], and symptoms of depression and suicide [38]. Chatbots were also used for self-management (n=7), counselling (n=5), education (n=4), and diagnosis (n=2).

Chatbots were implemented in stand-alone software in 70% of studies whereas the remaining chatbots were implemented in web-based platforms (Table 2). In the majority of studies (92.5%), chatbots generated their responses based on some predefined rules or decision trees (rule-based). In the remaining 4 studies, chatbots generated their responses based on machine learning approaches (Table 2). Specifically, while one study used supervised machine learning, another study used reinforcement learning. However, the remaining two studies did not specify the machine learning approaches used.

While chatbots led the dialogue in 86.8% of studies, both chatbots and users could lead the dialogue in the remaining 7 studies. Users could interact with the chatbots using: only written language via keyboards and mouse (n=26), only spoken language via microphones (n=8), a combination of spoken and visual languages via microphones and camera or Kinect (n=10), a combination of written and spoken languages (n=7), and a combination of written and visual languages (2). However, chatbots used the following modalities to interact with users: a combination of written (texts), spoken and visual languages (n=28), a combination of spoken and visual languages (n=11), only written language (n=9), and a combination of written and visual languages (n=5) (Table 2).

In 44 studies (83%), virtual representations (e.g. avatar or virtual human) were embodied in chatbots (Table 2). Chatbots targeted the following disorders: depression (n=16), autism (n=10), posttraumatic stress disorder (n=7), anxiety (n=7), substance use disorders (n=5), schizophrenia (n=3), dementia (n=3), stress (n=2), phobia (n=2), eating disorders (n=1), and any mental disorder (n=7). Appendix E shows characteristics of the chatbot used in each study.

Table 2: Features of chatbots in the included studies

Characteristics	Number of studies	Study ID[1]
Purpose[2]	Therapy:17	5, 9, 12-18, 22, 24-26, 36, 42, 47, 48
	Training:12	1, 6, 20, 21, 27, 35, 38-41, 44, 46
	Screening:10	2, 5, 10, 16, 23, 24, 28, 45, 49, 51
	Self-management:7	3, 4, 7, 8, 31, 32, 50
	Counselling:5	11, 33, 43, 52, 53
	Education:4	11, 19, 34, 37
	Diagnosing:2	29, 30
Platform	Stand-alone software:37	2-7, 10, 13, 17, 18, 20, 21, 23-32, 34, 38-41, 44-53
	Web-based:16	1, 8, 9, 11, 12, 14-16, 19, 22, 33, 35-37, 42, 43
Response generation	Rule-based:49	1-13, 15, 16, 18-42, 44-51, 53
	Artificial intelligence:4	14, 17, 43, 52
Dialogue initiative	Chatbot:46	1-15, 17-19, 22, 23, 25-30, 33-42, 44-53
	Both:7	16, 20, 21, 24, 31, 32, 43
	User:0	-
Input modality	Written:26	2-5, 7-9, 11, 12, 14-19, 25, 26, 29, 33, 34, 36, 37, 42, 43, 47, 48
	Spoken & Visual:10	1, 10, 21, 28, 35, 45, 46, 49-51
	Spoken:8	6, 13, 23, 31, 32, 44, 52, 53
	Written & Spoken:7	20, 24, 30, 38-41
	Written & Visual:2	22, 27
	Written, Spoken & Visual:28	1-3, 5, 7, 19-22, 24, 26, 27, 30, 33, 35-41, 43-49
	Spoken & Visual:11	6, 10, 13, 23, 28, 31, 32, 50-53

Output modality	Written:9	8, 9, 11, 12, 14, 16, 17, 25, 29
	Written & Visual: 5	4, 15, 18, 34, 42
Embodiment	Yes:44	1-7, 10, 13, 15, 18-24, 26-28, 30-53
	No:9	8, 9, 11, 12, 14, 16, 17, 25, 29
Targeted[3] disorders	Depression:16	3, 5, 7-10, 12, 14, 17, 24, 26, 30-33, 43
	Autism:10	1, 6, 19, 21, 27, 29, 35, 40, 44, 46
	Posttraumatic stress disorder:7	10, 23, 39, 43, 47, 48, 51
	Mental disorders:7	25, 36, 38, 41, 42, 50, 53
	Anxiety:7	8-10, 12, 14, 18, 24
	Substance use disorders:5	2, 11, 15, 22, 52
	Schizophrenia:3	4, 20, 34
	Dementia:3	28, 45, 49
	Phobia:2	13, 29
	Stress:2	8, 16
	Eating disorders:1	37
Tips	[1]: It is the number given for each included study as shown in Appendix C. [2]: Numbers do not add up as 4 chatbots had two different purposes. [3]: Numbers do not add up as several chatbots focused on more than one mental disorder.	

4 Discussion

4.1 Principal findings

This scoping review aimed to provide an overview of the features of chatbots used by individuals for their mental health as reported in the empirical literature. We identified 53 studies that assessed 41 different chatbots. The most common use of chatbots was delivery of therapy, training, and screening. Of 17 chatbots providing therapy, 10 chatbots were based on cognitive behavioural therapy, and 8 chatbots targeted people with depression and anxiety. Most chatbots (8 of 12) that provide training focused on people with autism. They also aimed to improve social skills (n=8) or job interviewing skills (n=4). Chatbots that were used as a screening tool focused mainly on depression (n=3), dementia (n=3), and posttraumatic stress disorder (n=3).

Most chatbots (70%) were implemented in stand-alone software. This was surprising because web-based chatbots are considered more appropriate than stand stand-alone software for the following two reasons. Firstly, to use web-based chatbots, users do not need to install a specific application to their devices, thereby reducing the risk of breaching their privacy. Secondly, web-based chatbots more accessible than stand-alone chatbots. This is because developers of a stand-alone chatbot need to create several applications of that chatbot in order for individuals to use it through different operating systems (e.g. Google's Android OS, Apple iOS, Apple macOS, Microsoft Windows, and Linux).

The majority of chatbots (92.5%) depended on only decision trees to generate their responses; only 7.5% used machine learning approaches. This may indicate that chatbots in mental health lag behind chatbots in other fields (e.g. customer services) where artificial intelligence chatbots are more common [39, 40]. Lack of artificial intelligence chatbots in health was also noted in a review conducted by Laranjo and colleagues [29]. The extensive use of rule-based approaches in the studies may be attributed to the fact that they are more appropriate for chatbots that perform simple, straightforward and well-structured tasks (e.g. information retrieval & data collection) [28, 29]. This was obvious in our review; all chatbots performing simple tasks (n=15) such as education, screening, and diagnoses were rule-based. In comparison with artificial intelligence chatbots, rule-based chatbots are more simple to develop and less prone to errors as their responses are predetermined and they do not need to create new responses [41]. Further, rule-based chatbots are considered more secure and responsible than artificial intelligence chatbots [42]. In contrast to artificial intelligence chatbots, users of rule-based chatbots cannot usually control the dialogue because their inputs are restricted to predefined words and phrases [28]. Further, rule-based chatbots cannot respond to users' inputs outside of the determined rules [41-43]. Although rule-based chatbots can be used for complex tasks and queries [42, 43], such chatbots are difficult to build because it is hard to expect every possible scenario and consumes too much time [41].

In 87% of the included studies, chatbots led and controlled the conversation. This was expected since most chatbots used rule-based approaches, which restrict user input to predefined words and replies, thereby, prevent them to lead or control the dialogue [28]. The rareness of chatbots led by users was also found in another review [29].

While the most common input modality was writing language only (49%), the most frequently used output modality was a combination of written, spoken and visual languages (53%). All chatbots using the written language as an input modality used the written language as an output modality too (n=34). The majority of chatbots in this review (83%) included virtual representations.

In our review, chatbots focused mainly on depression and autism. This may be attributed to the fact that depression and anxiety are the most common mental disorder around the world [44, 45], and chatbots are reported to be effective in improving social skills for patients with autism [14, 17, 46-48].

4.2 Strengths and limitations

4.2.1 Strengths

This review provides a list of chatbots in mental health that are classified according to their features, and it describes the main characteristics of previous studies in this field. This helps readers to explore how chatbots have been used in mental health and research activity in this field.

The current review is the first that focused on embodied and non-embodied chatbots used for any mental disorder. This made our review more comprehensive than the two previous reviews [6, 22]; where the former focused on some mental disorders and the latter focused on only embodied chatbots. Therefore, this comprehensive review provides a holistic view of the field and enables readers to explore more choices of chatbots for more varied mental disorders.

In comparison with the previous reviews [6, 22], the current review is the only review that searched Google Scholar and performed backward and forward reference list checking in order to identify grey literature. This enabled us to minimise the risk of publication bias. Further, the current review was the only one in the field that was developed, executed, and reported according to the PRISMA Extension for Scoping Reviews [23]. Accordingly, this produced a high quality scoping review.

The study selection process and data extraction process were conducted by two reviewers independently to reduce selection bias. Agreement between reviewers was very good for the study selection process and good for the data extraction process.

4.2.2 Limitations

Although chatbots were included in this review regardless of their purpose, type of response generation, input and output modalities, embodiment, and targeted disorders, we restricted their platforms to stand-alone software and web browser (but not robotics, serious games, SMS, nor telephones) and restricted their type of dialogue initiative to users and system (but not human operator). This led to excluding more than 25 studies. We excluded systems that are controlled by human operator as they are more similar to telemedicine systems than chatbots in terms of system design. The above-mentioned restriction was also applied in the previous two reviews [6, 22].

As the field of the current review is interdisciplinary, we searched several bibliographic databases from different fields (e.g. Medline & PsycINFO from the health field, and IEEE Xplore and ACM digital library from computer science field). Owing to practical constraints, we could not search other databases that cover both fields (e.g. Web of Science and ProQuest), conduct manual search, and contact experts. Accordingly, it is likely that we have missed some studies.

The current review excluded several papers as they were not primary studies. Therefore, this review may miss several chatbots reported in proposals, book chapters, and conference abstracts. Different taxonomies are available for classifying conversation agents. Therefore, we adapted several taxonomies [28-30] to characterise the chatbots in our review.

4.3 Practical and research implications

4.3.1 Practical implications

We have summarised here a list of different chatbots for different mental health disorders. Healthcare professionals can use this list to help potential users to identify the most appropriate chatbot for their mental health.

Relatively few chatbots in our review targeted patients with schizophrenia, dementia, phobic disorders, stress, and eating disorders. Also, no chatbots targeted patients with obsessive-compulsive disorder and bipolar. Therefore, chatbots' developers should develop more chatbots that target the above-mentioned disorders.

The current review identified relatively few chatbots implemented on a web-based platform despite its advantages mentioned in Section 4.1. We recommend chatbots' designers to consider using a web-based platform. Latest statistics showed that there are more than 4.437 billion internet users around the world in April 2019, with a growth rate of 8.6% over the last 12 months [49]. This indicates that web-based chatbots can be accessible to a vast number of users.

Although artificial intelligence chatbots can generate replies to complicated questions and enable users to lead the conversation, they were relatively few in mental health. Artificial intelligence chatbots depend on natural language processing to understand the context and intent of a question and to respond to it [42, 43]. Further, they mainly use machine and deep learning to learn from gathered data and continuously improve their performance and responses [42, 43]. Artificial intelligence chatbots also need large health databases and corpora in order for them to be trained [50]. Given the advancements in machine and deep learning and natural language processing and growing availability of large health databases and corpora [51, 52], developers should endeavour to build more artificial intelligence chatbots. Although artificial intelligence chatbots are prone to errors, such errors can be minimised and diminished by extensive training and more use [41, 42].

In our review, only 4 chatbots were implemented in developing countries. Developing countries have more shortage of mental health professionals than developed countries (0.1 per 1,000,000 people vs. 9 per 100,000 people) [7, 8], thereby, people in developing countries may be more in need of chatbots than those in developed countries. Developers should implement more chatbots in developing countries if the already implemented chatbots show their effectiveness.

4.3.2 Research implications

Our review showed that less than half of effectiveness studies (14 of 33) used randomised controlled trials. Given that randomised controlled trials are considered the most appropriate design for effectiveness studies [53, 54], researchers should conduct more randomised controlled trials to assess the effectiveness of chatbots in mental health. The sample size was very small (<50) in more than half of the studies, and many studies recruited non-clinical samples. We recommend future studies to recruit large clinical samples in order to increase the generalisability of the findings.

Few studies have assessed the adoption of chatbots, and none of the studies assessed the factors that affect users' adoption of them. It is well known that identifying factors that influence use of chatbots is very essential to improve their implementation success [55, 56]. Researchers should examine the factors that affect use of chatbots in mental health.

The included studies were inconsistent in how outcomes are measured (e.g. usability, severity of depression, and users' satisfaction). For example, while some studies assessed the usability of chatbots as a whole, other studies assessed the usability of specific components of chatbots (e.g. speech synthesis & dialogue management). Further, different usability questionnaires were used by those studies. To ease the interpretation of studies' results and compare between chatbots, researchers should standardise measures of the same outcomes.

The included studies were also inconsistent in reporting characteristics of chatbots, study methods, and characteristics of the sample. For example, it was challenging for the reviewers to identify chatbots' platforms, their input and output modality, type of response generation, and dialogue initiative. There are several reporting guidelines for researchers, such as the Consolidated Standards of Reporting Trials of electronic and mobile health applications and online telehealth (CONSORT-EHEALTH) [57] and the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement [58]. In addition that such guidelines improve reporting of studies, they enable readers to assess the quality of studies, combine their results, and compare between interventions [29]. Future research should be reported according to guidelines appropriate for its aim and study design. Further, researchers should enable readers to access the chatbot, or provide images of its main features.

The current scoping review did not summarise the results of included studies for two reasons. The first, this review aimed to identify the main features of chatbots in mental health and the research activity in this field, thereby, helping researchers in identifying and prioritising the main gap in the literature. The second, this review included a large number of studies, which assessed different outcomes, thereby, summarising their results for each outcome in one review is not practical. Hence, we encourage researchers to conduct systematic reviews to summarise the results of the studies for each outcome. Further, future systematic reviews should focus on acceptability and effectiveness of the chatbots because there are numerous studies assessed those outcomes (more than 33 studies) and, to the best of our knowledge, findings of those studies were not summarised by systematic reviews.

5 Conclusion

We identified 41 different chatbots in mental health. Most chatbots were rule-based, implemented in stand-alone software, initiated the dialogue, contained embodiment, focused on depression and autism, and implemented in developed countries. The most common use of chatbots was delivery of therapy, training, and screening. The most common input modality was writing language only, and the most frequently used output modality was a combination of written, spoken and visual languages.

Research regarding chatbots in mental health is emerging, and this was apparent from publication year of studies (most studies were published in the last 9 years), lack of randomized controlled trials, and the inconsistency of the studies in terms of outcome measures and reporting of characteristics of chatbots and study design. Future studies should be consistent in measuring the outcomes and follow published guidelines to standardise their reporting.

Authors’ contributions

Alaa Abd-alrazaq developed the protocol and conducted the search with guidance from and under the supervision of Mowafa Househ & Bridgette M Bewick. Study selection and data extraction were carried out independently by Alaa Abd-alrazaq & Mohannad Alajlani. Alaa Abd-alrazaq and Ali Alalwan drafted the manuscript, and it was revised critically for important intellectual content by all authors. All authors approved the manuscript for publication and agree to be accountable for all aspects of the work.

Acknowledgements

The authors would like to thank Jon Bandola and Zeineb Safi for his help in providing some thoughts about the introduction and data extraction.

Statement on conflicts of interest

The authors have no competing interests to declare.

Summary table

Summary table
What was already known on this topic: • Mental disorders are prevalent worldwide and are one of the most common causes of disability. • Chatbots may be useful tools for patients with mental disorders, especially those who are reluctant to seek mental health advice due to stigmatization. • Studies have been conducted to assess different aspects of using chatbots for mental health. **What this study added to our knowledge:** • There are numerous chatbots that are used for various mental disorders and purposes, have different platforms, types of response generation, types of dialogue initiative, input modalities, and output modalities. • Studies are inconsistent in how outcomes are measured, reporting of study design and of characteristics of chatbots.

References

[1] World Health Organisation, Mental health: a state of well-being, 2014.
[2] Z. Steel, C. Marnane, C. Iranpour, T. Chey, J.W. Jackson, V. Patel, D. Silove, The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013, Int J Epidemiol, 43 (2014) 476-493.
[3] World Health Organization, The global burden of disease: 2004 update, World Health Organization, Geneva, 2008.
[4] H.A. Whiteford, A.J. Ferrari, L. Degenhardt, V. Feigin, T. Vos, The global burden of mental, neurological and substance use disorders: an analysis from the Global Burden of Disease Study 2010, PloS one, 10 (2015) e0116820.
[5] S.P. Jones, V. Patel, S. Saxena, N. Radcliffe, S. Ali Al-Marri, A. Darzi, How Google's 'ten things we know to be true'could guide the development of mental health mobile apps, Health Affairs, 33 (2014) 1603-1611.
[6] A.N. Vaidyam, H. Wisniewski, J.D. Halamka, M.S. Kashavan, J.B. Torous, Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape, The Canadian Journal of Psychiatry, (2019) 0706743719828977.
[7] C.J. Murray, T. Vos, R. Lozano, M. Naghavi, A.D. Flaxman, C. Michaud, M. Ezzati, K. Shibuya, J.A. Salomon, S. Abdalla, Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010, The lancet, 380 (2012) 2197-2223.
[8] B.D. Oladeji, O. Gureje, Brain drain: a challenge to global mental health, BJPsych International, 13 (2016) 61-63.
[9] E. Anthes, Mental health: there's an app for that, Nature News, 532 (2016) 20.
[10] R.D. Hester, Lack of access to mental health services contributing to the high suicide rates among veterans, International journal of mental health systems, 11 (2017) 47.
[11] V.M. Kumar, A. Keerthana, M. Madhumitha, S. Valliammai, V. Vinithasri, Sanative Chatbot For Health Seekers, Int J Eng Comp Sci, 5 (2016) 16022-16025.
[12] G.M. Lucas, A. Rizzo, J. Gratch, S. Scherer, G. Stratou, J. Boberg, L.P. Morency, Reporting mental health symptoms: Breaking down barriers to care with virtual human interviewers, Front. Robot. AI, 4 (2017).
[13] B. Inkster, S. Sarda, V. Subramanian, An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study, JMIR Mhealth Uhealth, 6 (2018) e12106.
[14] M.R. Ali, Z. Rasazi, A.A. Mamun, R. Langevin, R. Rawassizadeh, L. Schubert, M.E. Hoque, A Virtual Conversational Agent for Teens with Autism: Experimental Results and Design Lessons, arXiv preprint arXiv:1811.03046, (2018).
[15] J. Huang, Q. Li, Y. Xue, T. Cheng, S. Xu, J. Jia, L. Feng, Teenchat: a chatterbot system for sensing and releasing adolescents' stress, International Conference on Health Information Science, Springer, 2015, pp. 133-145.
[16] M.D. Pinto, R.L. Hickman, Jr., J. Clochesy, M. Buchner, Avatar-based depression self-management technology: promising approach to improve depressive symptoms among young adults, Applied Nursing Research, 26 (2013) 45-48.
[17] S.Z. Razavi, M.R. Ali, T.H. Smith, L.K. Schubert, M.E. Hoque, The LISSA virtual human and ASD teens: An overview of initial experiments, International Conference on Intelligent Virtual Agents, Springer, 2016, pp. 460-463.
[18] U. Lahiri, E. Bekele, E. Dohrmann, Z. Warren, N. Sarkar, Design of a virtual reality based adaptive response technology for children with autism, IEEE Trans Neural Syst Rehabil Eng, 21 (2013) 55-64.

[19] U. Yasavur, C. Lisetti, N. Rishe, Let's talk! speaking virtual counselor offers you a brief intervention, J. Multimodal User Interfaces, 8 (2014) 381-398.
[20] T.W. Bickmore, K. Puskar, E.A. Schlenk, L.M. Pfeifer, S.M. Sereika, Maintaining reality: Relational agents for antipsychotic medication adherence, Interacting with Computers, 22 (2010) 276-288.
[21] M.H. Luerssen, T. Hawke, Virtual Agents as a Service: Applications in Healthcare, Proceedings of the 18th International Conference on Intelligent Virtual Agents, ACM, Sydney, NSW, Australia, 2018.
[22] S. Provoost, H.M. Lau, J. Ruwaard, H. Riper, Embodied conversational agents in clinical psychology: a scoping review, Journal of Medical Internet Research, 19 (2017) e151.
[23] A.C. Tricco, E. Lillie, W. Zarin, K.K. O'Brien, H. Colquhoun, D. Levac, D. Moher, M.D. Peters, T. Horsley, L. Weeks, PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation, Annals of internal medicine, 169 (2018) 467-473.
[24] J. Higgins, J. Deeks, Chapter 7: selecting studies and collecting data, in: J. Higgins, S. Green (Eds.) Cochrane Handbook for Systematic Reviews of Interventions, John Wiley & Sons, Sussex, UK, 2008, pp. 151-185.
[25] D.G. Altman, Practical statistics for medical research, CRC press1990.
[26] H. Arksey, L. O'Malley, Scoping studies: towards a methodological framework, International journal of social research methodology, 8 (2005) 19-32.
[27] M.J. Grant, A. Booth, A typology of reviews: an analysis of 14 review types and associated methodologies, Health Information & Libraries Journal, 26 (2009) 91-108.
[28] M.F. McTear, Spoken dialogue technology: enabling the conversational user interface, ACM Computing Surveys (CSUR), 34 (2002) 90-169.
[29] L. Laranjo, A.G. Dunn, H.L. Tong, A.B. Kocaballi, J. Chen, R. Bashir, D. Surian, B. Gallego, F. Magrabi, A.Y. Lau, Conversational agents in healthcare: a systematic review, Journal of the American Medical Informatics Association, 25 (2018) 1248-1258.
[30] J. Chu-Carroll, M.K. Brown, Tracking initiative in collaborative dialogue interactions, Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, Association for Computational Linguistics, 1997, pp. 262-270.
[31] K.K. Fitzpatrick, A. Darcy, M. Vierhile, Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial, JMIR Ment Health, 4 (2017) e19.
[32] J. Martínez-Miranda, A. Bresó, J.M. García-Gómez, Look on the bright side: a model of cognitive change in virtual agents, International Conference on Intelligent Virtual Agents, Springer, 2014, pp. 285-294.
[33] M.J. Smith, E.J. Ginger, K. Wright, M.A. Wright, J.L. Taylor, L.B. Humm, D.E. Olsen, M.D. Bell, M.F. Fleming, Virtual reality job interview training in adults with autism spectrum disorder, Journal of Autism and Developmental Disorders, 44 (2014) 2450-2463.
[34] B. Mirheidari, Detecting early signs of dementia in conversation, University of Sheffield, 2018.
[35] H. Tanaka, H. Adachi, N. Ukita, M. Ikeda, H. Kazui, T. Kudo, S. Nakamura, Detecting Dementia Through Interactive Computer Avatars, IEEE Journal of Translational Engineering in Health and Medicine, 5 (2017) 8038769.
[36] T. Ujiro, H. Tanaka, H. Adachi, H. Kazui, M. Ikeda, T. Kudo, S. Nakamura, Detection of dementia from responses to atypical questions asked by embodied conversational agents, Proc. Interspeech 2018, (2018) 1691-1695.
[37] M. Auriacombe, S. Moriceau, F. Serre, C. Denis, J.-A. Micoulaud-Franchi, E. de Sevin, E. Bonhomme, S. Bioulac, M. Fatseas, P. Philip, Development and validation of a virtual agent to screen tobacco and alcohol use disorders, Drug and Alcohol Dependence, 193 (2018) 1-6.

[38] A. Breso, J. Martinez-Miranda, C. Botella, R. Banos, J. Garcia-Gomez, Usability and acceptability assessment of an empathic virtual agent to prevent major depression, Expert Systems: International Journal of Knowledge Engineering and Neural Networks, 33 (2016) 297-312.
[39] S. Mallios, N. Bourbakis, A survey on human machine dialogue systems, 2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA), 2016, pp. 1-7.
[40] I.V. Serban, R. Lowe, P. Henderson, L. Charlin, J. Pineau, A survey of available corpora for building data-driven dialogue systems, arXiv preprint arXiv:1512.05742, (2015).
[41] K. Shridhar, Rule based bots vs AI bots, Medium.com, 2017.
[42] Hubtype, Rule-Based vs AI Chatbots, 2018.
[43] A. Dua, Rule-Based Bots or AI Bots? Which One Do You Prefer?, 2019.
[44] A.J. Ferrari, F.J. Charlson, R.E. Norman, S.B. Patten, G. Freedman, C.J. Murray, T. Vos, H.A. Whiteford, Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010, PLoS medicine, 10 (2013) e1001547.
[45] H. Ritchie, M. Roser, Mental Health, Our World in Data, 2018.
[46] M. Milne, M.H. Luerssen, T.W. Lewis, R.E. Leibbrandt, D.M.W. Powers, Development of a virtual agent based social tutor for children with autism spectrum disorders, The 2010 International Joint Conference on Neural Networks (IJCNN), 2010, pp. 1-9.
[47] H. Tanaka, H. Negoro, H. Iwasaka, S. Nakamura, Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders, PLoS ONE Vol 12(8), 2017, ArtID e0182151, 12 (2017).
[48] H. Tanaka, S. Sakti, G. Neubig, T. Toda, H. Negoro, H. Iwasaka, S. Nakamura, Automated Social Skills Trainer, Proceedings of the 20th International Conference on Intelligent User Interfaces - IUI '15, 2015, pp. 17-27.
[49] Datareportal, Digital 2019: Q2 Global Digital Statshot, 2019.
[50] R. López-Cózar, Z. Callejas, G. Espejo, D. Griol, Enhancement of conversational agents by means of multimodal interaction, in: D. Perez-Martin, I. Pascual-Nieto (Eds.) Conversational Agents and Natural Language Interaction: Techniques and Effective Practices, IGI Global2011, pp. 223-252.
[51] T.B. Murdoch, A.S. Detsky, The inevitable application of big data to health care, Jama, 309 (2013) 1351-1352.
[52] N.M. Radziwill, M.C. Benton, Evaluating quality of chatbots and intelligent conversational agents, arXiv preprint arXiv:1704.04579, (2017).
[53] A. Bhattacherjee, Social science research: Principles, methods, and practices, 2012.
[54] B. Matthews, L. Ross, Research methods: a practical guide for the social sciences, Pearson Education, New York, USA, 2010.
[55] M.W. Huygens, J. Vermeulen, R.D. Friele, O.C. van Schayck, J.D. de Jong, L.P. de Witte, Internet services for communicating with the general practice: barely noticed and used by patients, Interactive journal of medical research, 4 (2015) e21.
[56] V. Fung, E. Ortiz, J. Huang, B. Fireman, R. Miller, J.V. Selby, J. Hsu, Early experiences with e-health services (1999-2002): promise, reality, and implications, Medical care, (2006) 491-496.
[57] G. Eysenbach, C.-E. Group, CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions, Journal of Medical Internet Research, 13 (2011) e126.
[58] E. Von Elm, D.G. Altman, M. Egger, S.J. Pocock, P.C. Gøtzsche, J.P. Vandenbroucke, The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies, Annals of internal medicine, 147 (2007) 573-577.