

Review

# Perceptions and Opinions of Patients About Mental Health Chatbots: Scoping Review

Alaa A Abd-Alrazaq<sup>1</sup>, BSc, MSc, PhD; Mohannad Alajlani<sup>2</sup>, BSc, MSc, MBA, PhD; Nashva Ali<sup>1</sup>, BSc, MSc; Kerstin Denecke<sup>3</sup>, Dr rer nat; Bridgette M Bewick<sup>4</sup>, BA, MA, PhD; Mowafa Househ<sup>1</sup>, BCom, MEng, PhD

<sup>1</sup>Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

<sup>2</sup>Institute of Digital Healthcare, University of Warwick, Warwick, United Kingdom

<sup>3</sup>Institute for Medical Informatics, Bern University of Applied Science, Bern, Switzerland

<sup>4</sup>Leeds Institute of Health Sciences, School of Medicine, University of Leeds, Leeds, United Kingdom

**Corresponding Author:**

Mowafa Househ, BCom, MEng, PhD

Division of Information and Computing Technology

College of Science and Engineering, Hamad Bin Khalifa University

Qatar Foundation

PO Box 5825, Doha Al Luqta St, Ar-Rayyan

Doha, RT542

Qatar

Phone: 974 55708549

Email: [mhouseh@hbku.edu.qa](mailto:mhouseh@hbku.edu.qa)

## Abstract

**Background:** Chatbots have been used in the last decade to improve access to mental health care services. Perceptions and opinions of patients influence the adoption of chatbots for health care. Many studies have been conducted to assess the perceptions and opinions of patients about mental health chatbots. To the best of our knowledge, there has been no review of the evidence surrounding perceptions and opinions of patients about mental health chatbots.

**Objective:** This study aims to conduct a scoping review of the perceptions and opinions of patients about chatbots for mental health.

**Methods:** The scoping review was carried out in line with the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) extension for scoping reviews guidelines. Studies were identified by searching 8 electronic databases (eg, MEDLINE and Embase) in addition to conducting backward and forward reference list checking of the included studies and relevant reviews. In total, 2 reviewers independently selected studies and extracted data from the included studies. Data were synthesized using thematic analysis.

**Results:** Of 1072 citations retrieved, 37 unique studies were included in the review. The thematic analysis generated 10 themes from the findings of the studies: usefulness, ease of use, responsiveness, understandability, acceptability, attractiveness, trustworthiness, enjoyability, content, and comparisons.

**Conclusions:** The results demonstrated overall positive perceptions and opinions of patients about chatbots for mental health. Important issues to be addressed in the future are the linguistic capabilities of the chatbots: they have to be able to deal adequately with unexpected user input, provide high-quality responses, and have to show high variability in responses. To be useful for clinical practice, we have to find ways to harmonize chatbot content with individual treatment recommendations, that is, a personalization of chatbot conversations is required.

(*J Med Internet Res* 2021;23(1):e17828) doi: [10.2196/17828](https://doi.org/10.2196/17828)

**KEYWORDS**

chatbots; conversational agents; mental health; mental disorders; perceptions; opinions; mobile phone

## Introduction

### Background

Mental disorders are a growing global concern. Approximately 29% of individuals may experience such disorders in their lifetime [1]. Mental disorders have a negative effect on the quality of life and are one of the most common causes of disability [2]. Globally, the lost labor and capital output caused by mental disorders are predicted to cost US \$16 trillion between 2011 and 2030 [3]. For many, accessing mental health care services is challenging because of the shortage of mental health care providers [4-7], lack of health care insurance coverage [8], and perceived stigma [9-11]. The lack of access to mental health care services increases the risk of self-harm and suicide [12,13].

Technological advancements have improved access to mental health care services [3]. According to the World Health Organization, more than one-fourth of 15,000 mobile health (mHealth) apps focus on mental health diagnosis or support [13]. Chatbots, also called conversational agents, virtual agents, and dialog systems, are one of the main mobile apps used for mental health.

Chatbots are programs able to converse and interact with a human using voice, text, and animation [14]. There are 2 types of chatbots: rule-based chatbots and intelligent chatbots. Although the former uses some predefined rules or decision trees to manage its response and dialog, the latter uses artificial intelligence (AI) to generate its dialog [14]. Chatbots have been employed in businesses across different industries, such as e-commerce and retail (eg, eBay's ShopBot), hospitality (eg, Chatobook), real estate (eg, Apartment Ocean), entertainment (eg, Mojihunt), news (CNN's Chatbot), travel (eg, Hello Hipmunk), finance and banking (eg, Wells Fargo's chatbot), insurance (eg, ABIE), education (eg, Feed.Mind), legal (eg, Immigration Virtual Assistant), and personal assistance (eg, Apple's Siri). In addition to the abovementioned industries, chatbots have become popular in health care (in general) and mental health (in specific) in the past 5 years. According to a scoping review conducted by Abd-alrazaq et al [14], there were 41 different chatbots used for several purposes in mental health, such as therapy, training, education, counseling, and screening. A systematic review of 12 studies showed that chatbots are effective in improving some mental disorders, such as depression, stress, and acrophobia [15].

### Research Problem and Aim

The adoption of new technology relies on the perceptions and opinions of users. Numerous studies have been conducted to assess the perceptions and opinions of patients about mental health chatbots [14]. There is a need to explore perceptions and opinions that may help in predicting the adoption of chatbots and improving them [14]. Although several reviews have been conducted on chatbots in mental health [4,14-17], none have summarized the findings of previous studies regarding perceptions and opinions of patients about mental health chatbots. Accordingly, the aim of this study is to review the perceptions and opinions of patients about mental health chatbots, as reported in the previous literature.

## Methods

### Study Design

We conducted a scoping review to accomplish this objective. A scoping review was conducted as the aim was to map the body of literature on this topic [18]. Owing to the broad nature of the inquiry, we expected a diversity of study designs. Scoping reviews are more suited to broader aims [18]. As we were not seeking to summarize the best available research on a specific question, a systematic review was not the ideal choice. Using chatbots for mental health is a field in relative infancy; therefore, a broader aim was necessary. The range of study designs currently used in the field makes equitable risk of bias assessment difficult; it is acknowledged that the risk of bias assessments is not required in scoping reviews [18,19]. Scoping reviews are generally accepted as more appropriate when diversity of study designs is expected. The PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) Extension for Scoping Reviews guidelines were followed to carry out a systematic and transparent review [20].

### Search Strategy

#### Search Sources

The following electronic databases were searched in the current review: MEDLINE (via Ovid), Embase (via Ovid), PsycINFO (via Ovid), Scopus, Cochrane Central Register of Controlled Trials, IEEE Xplore, ACM Digital Library, and Google Scholar. Given that Google Scholar usually finds several thousands of references, which are ordered by their relevance to the search topic, we screened only the first 100 references [14,15,21]; these references are the most relevant. The search was conducted from October 25 to October 28, 2019. We also conducted backward reference list checking, where reference lists of the included studies and reviews on the search topic were screened for additional studies of relevance to the review. In addition, we carried out forward reference list checking, where the *cited by* function available in Google Scholar was used to find and screen studies that cited the included studies.

#### Search Terms

To derive search terms, we checked previous literature reviews [4,14-17] and opinions of informatics experts interested in mental health (which were collected in informal meetings). The search terms were selected based on the target intervention (eg, chatbots and conversational agents) and condition (eg, mental disorder and anxiety). The controlled vocabulary search terms (eg, MeSH, Emtree) were used, as appropriate. The search strings used for searching each electronic database are detailed in [Multimedia Appendix 1](#).

#### Study Eligibility Criteria

The intervention of interest in this review was chatbots that operate as stand-alone software or a web browser ([Textboxes 1 and 2](#)). We excluded chatbots that were integrated into robotics, serious games, SMS, or telephone systems and those that depended on human operator-generated dialog. No restrictions were applied regarding the type of dialog initiative (ie, use, system, mixed) and input and output modality (ie,

spoken, visual, and written). The eligible population included individuals who used chatbots to improve their psychological well-being or mental disorders but not physicians or caregivers who use chatbots for their patients. To be included, studies had to assess patients' perceptions and opinions about chatbots for mental health. The review included peer-reviewed articles,

dissertations, and conference proceedings, and it excluded reviews, proposals, editorials, and conference abstracts. Only studies written in English were included in this review. No restrictions were applied regarding the study design, study setting, comparator, year of publication, or country of publication.

#### Textbox 1. Inclusion criteria.

- Intervention: chatbots operate as stand-alone software or a web browser
- Population: patients who use chatbots for improving their psychological well-being or mental disorders
- Outcome: patients' perceptions and opinions about mental health chatbots
- Type of publication: peer-reviewed articles, dissertations, and conference proceedings
- Language: English

#### Textbox 2. Exclusion criteria.

- Intervention: chatbots integrated into robotics, serious games, SMS, or telephone systems and those depend on human operator-generated dialog
- Population: physicians or caregivers who use chatbots for improving their psychological well-being or mental disorders
- Outcome: other outcomes
- Type of publication: reviews, proposals, editorials, and conference abstracts
- Language: other languages

## Study Selection

In this review, MA and NA independently screened the titles and abstracts of all retrieved studies and independently read the full texts of studies included from the first step. AA resolved any disagreements between the reviewers. Cohen kappa was calculated to assess the intercoder agreement [22], which was 0.86 and 0.90 for screening titles and abstracts and for reading full texts, respectively, indicating excellent agreement [23].

## Data Extraction

Multimedia Appendix 2 shows the data extraction form used in this review, which was pilot tested using 4 included studies. Data were extracted from the included studies by 2 reviewers independently (MA and NA), and the third reviewer (AA) resolved any discrepancies in decisions between the reviewers. Cohen kappa showed an excellent agreement (0.83) [23].

## Assessment of Risk of Bias

Scoping reviews do not usually assess the risk of bias of the included studies because they have broad aims and include studies with diverse study designs [18,19]. The aim of this review was very broad, and the included studies had different study designs. Thus, the risk of bias of the included studies was not assessed in this review.

## Data Synthesis

A narrative approach was used to synthesize the data extracted from the included studies. Thematic analysis was used to generate themes based on the findings of the included studies.

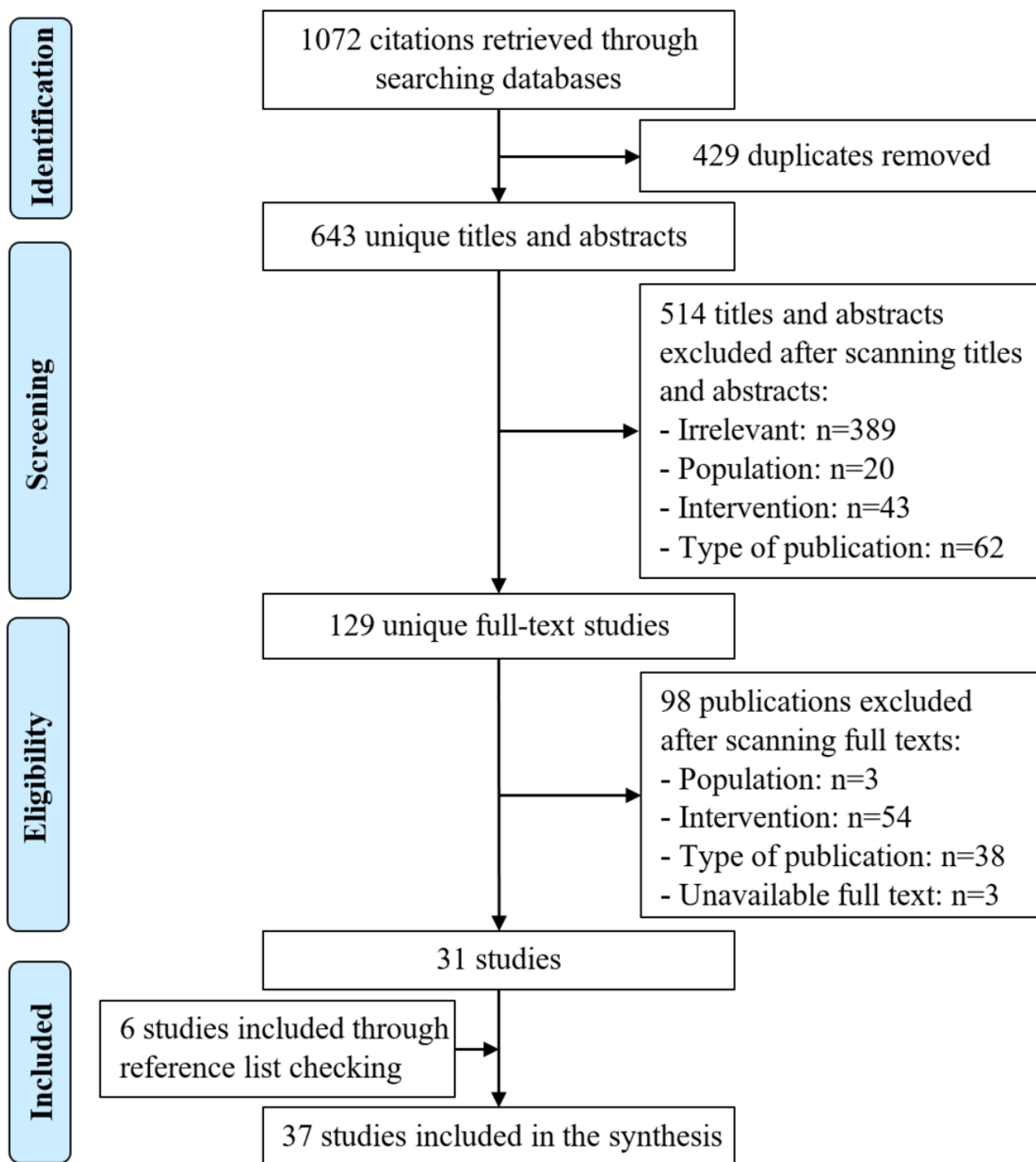
This data synthesis approach (ie, thematic analysis) has been applied in numerous systematic and scoping reviews [24-28]. Given the exploratory nature of this review, an inductive approach was used to generate themes directly from the data [29]. The thematic analysis was conducted following the steps proposed by Braun and Clarke [30], as it is the most systematic guide for conducting thematic analysis to date [31,32]. Specifically, the first author (AA) scrutinized and rescrutinized the extracted data to familiarize himself with it. AA then coded the data systematically. On the basis of the assigned codes, themes were generated. All authors checked the fit of the generated themes to the extracted data and assigned codes. Where authors had differing opinions on the assigned codes and/or generated themes, these discrepancies were resolved through discussion. Finally, all authors participated in defining and naming the themes. Microsoft Excel was used to manage the analysis process.

## Results

### Search Results

As shown in Figure 1, 1072 citations were found by searching the electronic databases. After removing 429 duplicates of these citations, 643 titles and abstracts were screened. In the screening process, we excluded 514 titles and abstracts. Reading the full text of the remaining 129 citations resulted in a further 98 citations being excluded. The reasons for the exclusion are detailed in Figure 1. In backward and forward reference checking, we found 6 additional studies. In total, 37 studies were included in this review.

Figure 1. Flowchart of the study selection process.



### Characteristics of Included Studies

As shown in Table 1, the most commonly used study design was a cross-sectional survey (34/37, 92%). About two-thirds of the studies were published as journal articles (24/37, 65%).

The included studies were conducted in more than 15 countries, but approximately 46% (17/37) of them were carried out in the United States. Approximately 62% (23/37) of the studies were published between 2015 and 2019.

**Table 1.** Characteristics of the included studies.

Parameters and characteristics	Studies <sup>a</sup>
<b>Study metadata, n (%)</b>	
<b>Study design</b>	
Survey	34 (92)
Quasi-experiment	2 (5)
Randomized controlled trial	1 (3)
<b>Type of publication</b>	
Journal article	24 (65)
Conference proceeding	12 (32)
Thesis	1 (3)
<b>Country</b>	
United States	17 (46)
Australia	3 (8)
France	3 (8)
The Netherlands	3 (8)
Japan	2 (5)
Germany	1 (3)
Korea	1 (3)
Spain	1 (3)
Sweden	1 (3)
Turkey	1 (3)
United Kingdom	1 (3)
Romania, Spain, and Scotland	1 (3)
Spain and Mexico	1 (3)
Global population	1 (3)
<b>Year of publication</b>	
Before 2010	3 (8)
2010-2014	11 (30)
2015-2019	23 (62)
<b>Population characteristics</b>	
<b>Sample size, n (%)</b>	
≤50	24 (65)
51-100	5 (14)
101-200	6 (16)
>200	2 (5)
<b>Age (years)</b>	
Mean (range) <sup>b</sup>	33.4 (13-79)
<b>Sex, n (%)</b>	
Male <sup>c</sup>	1436 (50)
<b>Sample type, n (%)</b>	
Clinical sample	21 (57)
Nonclinical sample	16 (43)

Parameters and characteristics	Studies <sup>a</sup>
<b>Setting, n (%)<sup>d</sup></b>	
Clinical	14 (38)
Educational	12 (32)
Community	8 (22)
<b>Intervention characteristics, n (%)</b>	
<b>Purpose</b>	
Therapy	12 (32)
Training	9 (24)
Self-management	6 (16)
Counseling	5 (14)
Screening	4 (11)
Diagnosing	1 (3)
<b>Platform</b>	
Stand-alone software	24 (65)
Web based	13 (35)
<b>Response generation</b>	
Rule based	32 (86)
Artificial intelligence	5 (14)
<b>Dialog initiative</b>	
Chatbot	32 (86)
Both	5 (14)
<b>Embodiment</b>	
Yes	30 (81)
No	7 (19)
<b>Targeted disorders<sup>e</sup></b>	
Depression	41 (23)
Autism	6 (16)
Anxiety	6 (16)
Any mental disorder	6 (16)
Substance use disorder	5 (14)
Posttraumatic stress disorder	5 (14)
Schizophrenia	3 (8)
Stress	3 (8)

<sup>a</sup>Percentages were rounded and may not sum to 100.

<sup>b</sup>Mean age was reported in 24 studies.

<sup>c</sup>Sex was reported in 29 studies.

<sup>d</sup>Setting was reported in 34 studies.

<sup>e</sup>Numbers do not add up as several chatbots target more than one health condition.

The sample size was 50 or less in 24 studies and more than 200 in 2 studies (Table 1). The participants' age was reported in 24 studies, with a mean age of participants was 33.4 years (SD 15.2; range 13-79 years). The sex of participants was reported in 29 studies, where the mean percentage of men was 49.5%. In 57% (21/37) of the studies, participants were from clinical

populations (ie, they had mental disorders). The sample was recruited from clinical (n=14), educational (n=12), or community settings (n=8). Multimedia Appendix 3 shows the metadata and population characteristics of each included study.

The 37 included studies assessed patients' perceptions and opinions about 32 different chatbots. Chatbots were used for therapeutic purposes (n=12), training (n=9), self-management (n=6) counseling (n=5), screening (n=4), and diagnosis (n=1; [Table 1](#)). Chatbots were implemented in stand-alone software in 65% (24/37) of the studies, whereas the remaining chatbots were implemented in web-based platforms. In the majority of studies (32/37, 86.5%), chatbots generated their responses based on predefined rules or decision trees (rule based). Chatbots in the remaining studies generated their responses based on machine learning approaches. Chatbots led the dialog in most studies (n=32/37, 86.5%); both chatbots and users could lead the dialog in the remaining studies. Virtual agents (eg, avatar or virtual human) were embodied in chatbots in 30 of 37 studies (81.1%). The most common disorder targeted by chatbots was depression (n=15, 40.5%). [Multimedia Appendix 4](#) shows the characteristics of the intervention in each included study [33-69].

## Study Findings

The thematic analysis generated 10 themes from the findings of the studies: usefulness, ease of use, responsiveness, understandability, acceptability, attractiveness, trustworthiness, enjoyability, content, and comparisons. More details about these themes are elaborated in the following subsections.

### Theme 1: Usefulness

In total, 20 studies investigated the usefulness of chatbots and/or their features for patients [33-52]. In 3 studies [41,47-49,51], the overall usefulness of chatbots was rated as high. Participants reported that chatbots are useful for practicing conversations in a private place [33,46], learning [37,38,42,46], making users feel better [40], preparing users for interactions with health care providers [43], implementing the learned skills in daily life [46], facilitating a sense of accountability from daily check-in [37,38], keeping the learned skills more prominently in users' minds [46], increasing users' self-efficacy [46], improving users' self-confidence and readiness for job interviews [47-49], and recalling users' memories [51]. However, participants in one study doubted the usefulness of chatbots for smoking cessation [39].

Users considered the following components of chatbots useful: real-time feedback [33,45,50], diary [52], weekly summary [42], presenting the helpline during the conversation [36], and psychoeducation [52]. Some studies found that virtual agents embodied in chatbots are useful for motivating users to use chatbots [34], establishing a relationship with them [35], and feeling supported by them [45]. However, other studies demonstrated that participants had neutral perceptions and opinions about the added value of embodied virtual agents with speech [52] or without speech [44].

### Theme 2: Ease of Use

The ease of use and usability of chatbots were assessed in 20 studies [33,34,36,39,43,46-51,53-61]. Participants in 15 studies rated the overall ease of use of chatbots as high [36,39,43,47-50,53-55,57-61]. A total of 5 studies assessed the usability of chatbots [34,36,46,51,56], and it was rated high in all these studies (ranging from 69 to 88.2). Participants in 3 studies reported that chatbots are easy to learn and become

familiar with [33,39,55]. Although participants did not find chatbots difficult to navigate in one study [33], they rated the navigation of the chatbot as low in another study [36].

In 3 studies, participants faced difficulty in using the chatbot because they did not know when [60] and how [61] to reply to it, and they did not have enough options to reply to it [62]. Some participants in 3 studies criticized chatbots because of technical glitches that happened during their use, such as screen freezing [37,45,60]. Users suggested that chatbots allow them to respond by speaking and not typing to make them easy to use [57].

### Theme 3: Responsiveness

This theme brings together perceptions and opinions of participants about verbal and nonverbal responses generated by chatbots in terms of realism, repetitiveness (variability), speed, friendliness, and empathy. A total of 10 studies assessed participants' perceptions and opinions about how real the chatbots were in terms of verbal and nonverbal responses. Although participants in 7 studies had mixed or neutral perceptions and opinions about the realism of verbal and nonverbal responses [33,35,38,45,46,50,56], most participants in 3 studies held corresponding positive perceptions and opinions [52,57,60]. Participants believed that chatbots may be more realistic if they have an embodied virtual agent [44] and are able to communicate verbally [33].

Most participants in several studies stated that chatbots were able to show friendly [34,52,59,60,63] and emotional [35,37,38] responsiveness. However, participants in other studies had mixed perceptions and opinions about the ability of chatbots to generate friendly [35,44,64] and emotional [64] responses. Participants in one study found chatbots with an embodied virtual agent friendlier than those without an embodied virtual agent [44].

A total of 7 studies demonstrated that chatbot responses were repetitive [35-38,40,42,61]. Participants in one study reported that the repetitiveness of responses made the chatbot look less real [42]. Moreover, 3 studies concluded that the speed of chatbot responses was appropriate [57,60,61]. Yet, participants in 2 studies criticized the speed of chatbot responses as either too fast [38] or too slow [33]. In 6 studies, participants did not face any difficulties in understanding and interpreting chatbot responses [45,57,59,60,65,66].

In general, participants in 6 studies were satisfied with chatbot responses [33-36,62,63,67]. Participants attributed their satisfaction to the accuracy and consistency of chatbot feedback with what their therapist recommended in the past [33], appropriate use of high-quality elements (eg, Graphics Interchange Format images) [36], consistency of voice and tone of the chatbot [36], and quality of information provided [62,63]. However, participants in 4 studies were not satisfied by the conversation of chatbot because of confusing responses [57], disturbing users [40], the robotic quality of its voice [62], using emoticons (emojis) [37], conversations being too short [37], and shallowness of the conversations [42].

Participants suggested several enhancements related to the responsiveness of chatbots, such as the ability to speak [44], more flowing speech [33], friendlier voice [44], immediate responses [33,45], faster blinks [33], more explanation [33],

providing feedback on each answered question [44], and more precise feedback [45].

#### **Theme 4: Understandability**

This theme brings together perceptions and opinions of participants about the ability of chatbots to understand their verbal and nonverbal contact. Chatbot understandability for verbal responses was rated as high among participants in 3 studies [33,45,61]; however, participants in other studies criticized the inability of the chatbot to understand their verbal responses in general [36-38,40,57], misspelled words (eg, anious instead of anxious) [36], different languages [36], unexpected answers [37,38], and nonverbal responses (eg, images, emojis, and facial expressions) [33,36,56]. Therefore, participants in 3 studies suggested that future chatbots should better understand and recognize their verbal and nonverbal responses [33,37,43].

#### **Theme 5: Acceptability**

This theme concerns participants' acceptability of chatbots and its functionalities and their intentions to use them in the future. The acceptability of chatbots was rated high by users in 12 studies [34,37,38,43,45,46,53,54,57,61,66,68]. Wearing eye tracker goggles [62] or the inclusion of an embodied virtual agent [65] was acceptable for participants. There were mixed opinions about the acceptability of chatbots in one study [35]. Chatbots connected with a perception system (camera) for extracting data about user behavior was not acceptable for most participants in one study [60]. Users in one study indicated that the acceptability of chatbots could increase if the purpose of chatbots was clearly explained [33]. Note that the lack of clarity of the purpose of chatbots was highlighted by participants in 2 studies [33,42].

Furthermore, 6 studies demonstrated that people would like to use chatbots in the future [35,43,45,46,50,53,54,61,65], whereas participants in 2 studies were ambivalent about the future use of chatbots [33,39]. This ambivalence resulted from participants perceiving chatbots as a conversational tool for children [33]. Some participants reported that they probably would not use chatbots unless they popped up on their screens [33]. Although participants in one study showed high intention to use the chatbot in the future, they highlighted that it is inconvenient or inappropriate to use it every day [35].

#### **Theme 6: Attractiveness**

Participants in one study rated the attractiveness of a chatbot as low [57]. Furthermore, Demirci [55] found that perceptions and opinions of users about attractiveness changed from positive before using the chatbot to neutral after using it. Participants suggested improvements of the controls (eg, icons, combo boxes, buttons, and font size) [33,34] and the appearance of the embodied virtual agent [34] to obtain a more attractive graphical user interface.

#### **Theme 7: Trustworthiness**

This theme concerns participants' trust in chatbot. In 7 studies, participants believed that chatbots are trustworthy [34,44,46,52,54,57,63]. One study concluded that participants were satisfied with the anonymity, confidentiality, and objectivity of chatbots [63]. Most participants in the 2 studies

were comfortable to share and did share personal information with the chatbot [56,63].

#### **Theme 8: Enjoyability**

Participants in 9 studies considered using chatbots as enjoyable and fun [36,42-44,47-49,52,55,59]. However, one study found that perceptions and opinions of users about enjoyment changed from positive before using the chatbot to neutral after using it [55].

#### **Theme 9: Content**

This theme contains participants' opinions about the content of chatbots. In 6 studies, participants were satisfied with the contents of chatbots such as videos, games, topics, suggestions, and weekly graphs [34,37,38,42,43,52]. However, the content of chatbots was criticized by users because of its superficiality [33,38], irrelevancy [38], having too long videos [37], and having overwhelming information [57]. Participants in 3 studies appreciated the personalization feature in chatbots that allowed them to customize the session length and the gender and appearance of the virtual agent [35,57,60]. Participants suggested that chatbots should contain psychoeducation [35], more therapy-based activities [34], longer and more frequent intervention sessions [43], more customizable content [35,43], and interventions for other chronic health conditions [43].

#### **Theme 10: Comparisons**

This theme brings together participant perspectives about chatbots in comparison with other chatbots or traditional methods. Although most participants in one study preferred interacting with a chatbot rather than a human for their health care [53], participants in another 2 studies had mixed preferences about that [33,45]. In 2 studies [44,58], participants preferred using chatbots with an embodied virtual agent compared with chatbots without an embodied virtual agent.

Participants in one study preferred that chatbot provides real-time feedback on their nonverbal behavior rather than postsession feedback [33]. According to Morris et al [67], participants preferred the chatbot's responses drawn from an existing pool of web-based peer support data rather than those generated by the chatbot itself. Users preferred to use chatbots that provide information in an interactive fashion [43], are added on a device that they already own (eg, smartphones, tablets, computers) [60], and call them by their first name [60].

A chatbot without an embodied virtual agent (text-based chatbot) was compared with 2 chatbots with an embodied virtual agent (one reacts to the user with verbal and nonverbal empathic reactions, whereas the other did not) in another study [58]. The study found that the empathic chatbot was more acceptable, enjoyable, empathizing, understanding, nicer, sociable, trustworthy, realistic, private, anthropomorphic, animated, intelligent, socially influencing, friendlier, and safer than the nonempathic chatbot and the text-based chatbot [58]. Furthermore, the study demonstrated that participants are more willing to disclose information to the text-based chatbot than to empathic and nonempathic chatbots and a human counselor [58]. The study also found that participants were more willing



to use empathic chatbots than nonempathic chatbots and text-based chatbots [58].

One study compared AI chatbots with an individual or a chatbot controlled by the same individual (Wizard-of-Oz) [56]. The study found that the Wizard-of-Oz chatbot was rated by participants as more usable and listened better than the AI chatbot [56]. Furthermore, users of the Wizard-of-Oz chatbot felt greater rapport than users of the AI chatbot and, surprisingly, than those who were interviewed by humans [56]. However, there was no difference between users of the AI chatbot and those interviewed by a human in their ratings of rapport [56].

In another study [69], participants felt a greater rapport with the real expert than with a rule-based chatbot. Participants also indicated that the rule-based chatbot is less able to understand their responses and feelings than a real expert [69]. Furthermore, participants found the real expert more trustworthy than the rule-based chatbot [69]. Participants reported that the chatbot is more enjoyable and engaging than traditional methods of learning and practicing dialectical behavior therapy skills [46].

## Discussion

### Principal Findings

The main finding of this review is that there are features of chatbots that health care providers cannot deliver over a long period. These features have been identified as useful in mental health chatbots: real-time feedback, weekly summary, and continuous data collection in terms of a diary. Usefulness and ease of use are aspects of chatbots that have been studied most comprehensively in the analyzed papers. Overall, the usefulness of mental health chatbots is perceived as high by patients. According to these studies, patients find chatbot systems easy to use. Interactional enjoyment and perceived trust are significant mediators of chatbot interaction [70]. Although chatbots are perceived as useful and easy to use, participants of reported studies recognized the existing conversational limitations of those systems: conversations are perceived as shallow, confusing, or too short. This points to an important issue to be addressed in future mental health chatbot development (see the *Practical and Research Implications* section). The conversation quality still needs to be improved. In this context, chatbot quality in terms of responsiveness and variability of responses is an important issue. Currently, systems are rather restricted in the number of responses, which might be because of the early development stage of many chatbots, as has already been reported by Laranjo et al [71]. Another relevant aspect judged important is the quality of provided information and consistency with recommendations of treating physicians. The implications of these principal findings toward practice and research are described in the *Practical and Research Implications* section.

### Comparison With Existing Literature

This is the first review that summarizes perceptions and opinions of patients about mental health chatbots, as reported by previous studies. Palanica et al [72] assessed physicians' perceptions of health chatbots. They found that physicians see the benefits of those apps basically in support of patients in managing their

health and improving physical, psychological, and behavioral health outcomes. As limitations of health chatbots, they mentioned the inability to comprehend and express human emotions and a lack of medical expert knowledge and intelligence. Our results provide the counterpart of this observation, namely, patients recognizing limitations in the conversation quality of health chatbots. A recent paper on a chatbot for regulating emotions shows that perceptions of patients and psychologists differ regarding attractiveness, efficiency, and stimulations of health chatbots [73]. Although psychologists see these aspects rather positive, patients are more critical and complain about the restrictions of chatbot conversations.

In their review of the landscape of psychiatric chatbots, Vaidyam [4] identified studies that showed high satisfaction with the interventions they received. Participants reported the interventions as helpful, easy to use, and informative and rated satisfaction highly (>4.2 out of 5) on all scales, including ease of use, desire to continue using the system, liking, and trust. In addition, the results of Bendig et al [16] suggest the practicability, feasibility, and acceptance of using chatbots to promote mental health. Our results confirm these observations: patients consider chatbots as useful, and acceptability is rated high.

A study assessed the use of mobile technologies in health-related areas from various perspectives [74]. It points to the following risks for mHealth app usage: lack of functionality, dissemination of false information, misdiagnosis, mistreatment, and unknown unwanted side effects. From the patient perspective, these issues could also be identified in our review: quality of provided information and consistency with recommendations of treating physicians are relevant aspects. The results of the study by Albrecht [74] go beyond our results by pointing to the risks of mHealth apps in case of an emergency. Implemented algorithms still lack reactivity (eg, in the case of self-endangerment or hazards of others). In addition, Singh et al [75] showed that only 23% of mHealth apps responded adequately to dangerous user input (eg, suicidal ideations). This illustrates the enormous need for improvement in terms of responsiveness of mHealth apps in potentially dangerous situations. According to our results, the patients did not seem to be too concerned about this issue of chatbot behavior in emergencies. It was not mentioned in the reviewed studies.

### Practical and Research Implications

#### Practical Implications

The study results have the following practical implications. To be useful, we need to create high-quality chatbots that are able to respond to a user in multiple ways. A mental health chatbot must be empathic to be perceived as motivating and engaging and to establish a relationship with the user. A study by de Gennaro [76] supports this by demonstrating that empathic chatbots have the potential to provide emotional support to victims of social exclusion.

The patient-doctor or patient-therapist relationship in standard health care settings is characterized by trust and loyalty. Measurements must be undertaken to make the chatbot-patient

relationship also trustworthy. This could be realized by providing information on the secondary use of the collected patient data on data storage and analysis procedures. Another approach is blended therapy [77], a combination of face-to-face and web-based or digital therapy, which shows the potential of a cost-effective and accessible format in cognitive behavioral therapy. This would also address another practical implication, which is that the chatbot has to be related to the therapy. In particular, recommendations provided by a chatbot must be consistent with the recommendations of the treating health care professionals. This demands the integration of chatbots into the health care process; the chatbot should be aware of the recommendations or treatment plans of the health care provider. Finally, to increase the acceptance of chatbot use in patients, physicians need to be convinced of the usefulness of those systems so that they will recommend them to patients. Studies suggest that there are already physicians who are convinced of the usefulness [72]. Given the strong bond of trust among patients toward their physicians, patients will be convinced of the usefulness of an app once their physician recommends it.

### **Research Implications**

From the practical implications, we can derive the following research implications. There is still a need to improve the linguistic capabilities of mental health chatbots [71]. Their ability to understand and react appropriately to user input has to be increased. Furthermore, methods are required to generate dynamic answers to ensure the variability of chatbot responses. Linguistic or lexical variability can be added to the knowledge base of rule-based chatbots, but the capabilities will always depend on the completeness of the knowledge base. Methods for slightly adapting or reformulating responses from the knowledge base could help in addressing this issue. In domains outside the health care domain, crowdsourcing has been applied to improve conversation quality [78]. However, in health care, we have to be careful with learning from data because we have to ensure that responses and recommendations are in line with clinical evidence. It is still an open research question on how to learn clinical evidence to train health chatbots.

Furthermore, methods have to be developed to deal with unexpected user input and to detect critical situations. In mental health, it is crucial to react appropriately for people who are at risk of suicide or self-harm [79]. Sentiment analysis methods have proven to be successful in analyzing social media messages with respect to suicide and self-harm [80]. These methods could be useful in health chatbots as well. The main challenge is the use of an appropriate reaction once an emergency situation has been detected. Another interesting research topic is the customization or personalization of chatbots to individual users. This topic is still in its infancy [81]. As long as mental health chatbots rely on decision trees or fix implemented rule bases, they will not be able to adapt to specific user needs. We can construct the knowledge base in such a way that there are responses for different types of users, but this will be time consuming and will always be incomplete. Learning from conversations with users could help. The style or complexity of language can be adapted based on the given user input. Patient-specific knowledge, for example, on treatment plans could be retrieved from health records. Methods are required

to include such knowledge dynamically to a chatbot. In this way, the chatbot content is adapted to match individual needs.

For evaluating the mental health chatbot, benchmarks have to be created, and consistent metrics and methods have to be developed. Laranjo et al [71] reviewed the characteristics, current applications, and evaluation measures of health chatbots. Evaluation measures were divided into 3 main types: technical performance, user experience, and health research measures. The first attempts toward evaluation frameworks for digital health interventions [82] and health chatbots [83,84] have been recently published. Depending on the facet under consideration, different metrics can be used: system performance and effectiveness can be evaluated using different computational metrics (eg, usability, ease of use, usefulness) [85]. Software quality can be measured by reliability, security, maintainability, and efficiency using software engineering metrics [86]. If the system uses AI and machine learning techniques, the metrics comprise the accuracy and precision of predictions and recommendations. Furthermore, the efficiency of the systems has to be evaluated and compared with existing care models. With regard to safe app use, 3 criteria should be evaluated: (1) quality of the therapeutic content, (2) functionality, and (3) data safety and protection [87].

### **Strengths and Limitations**

#### **Strengths**

This review was developed, executed, and reported according to the PRISMA Extension for Scoping Reviews [20]. This enabled us to produce a high-quality review.

The most commonly used databases in health and information technology were searched to retrieve relevant studies as many as possible. Searching Google Scholar and carrying out backward and forward reference list checking enabled us to identify gray literature and minimize the risk of publication bias as much as possible. As no restrictions were applied regarding the study design, study setting, comparator, year of publication, and country of publication, this review can be considered comprehensive.

Selection bias in this review was minimal because study selection and data extraction were performed independently by 2 reviewers. Furthermore, the agreement between reviewers was very good for study selection and data extraction. This study is one of the few reviews that used thematic analysis to synthesize the findings of the included studies. The thematic analysis followed the highly recommended guidelines proposed by Braun and Clarke [30].

#### **Limitations**

This review focused on chatbots that only work on stand-alone software and a web browser (but not robotics, serious games, SMS, or telephones). Furthermore, this review was restricted to chatbots that are not controlled by human operators (Wizard-of-Oz). Therefore, perceptions and opinions of patients found in this review may be different from their perceptions and opinions about Wizard-of-Oz chatbots and/or chatbots with alternative modes of delivery. The abovementioned restrictions

were applied by previous reviews about chatbots, as these features are not part of ordinary chatbots [4,14,17].

Owing to practical constraints, we restricted the search to English studies and we could not search interdisciplinary databases (eg, Web of Science and ProQuest), conduct manual search, or contact experts. Consequently, it is likely that we have missed some English and non-English studies. Most included studies were conducted in developed countries, particularly in the United States. Therefore, the findings of this review may not be generalizable to developing countries, as patients in such countries may have different perceptions and opinions about mental health chatbots.

## Conclusions

In this paper, we explored perceptions and opinions of patients about mental health chatbots, as reported in the existing

literature. The results demonstrated that there are overall positive perceptions and opinions of patients about mental health chatbots, although there is some skepticism toward trustworthiness and usefulness. Many important aspects have been identified to be addressed in research and practice. Among them are the need to improve the linguistic capabilities of chatbots and seamless integration into the health care process. Future research will have to pick up those issues to create successful, well-perceived chatbot systems, and we will start developing corresponding concepts and methods. The research implications are also relevant for health care chatbots beyond mental health chatbots. Their consideration has the potential to improve patients' perceptions of health care chatbots in general.

---

## Acknowledgments

The publication of this study was funded by the Qatar National Library. This study was a part of a project funded by the Qatar National Research Fund (NPRP12S-0303-190204). The project title is *A Personalized and Intelligent Digital Mental Health Platform for Qatar and the Arab world*.

---

## Authors' Contributions

AA developed the protocol and conducted a search with guidance from and under the supervision of MH and BB. Study selection and data extraction were performed independently by MA and NA. AA executed the analysis, and all authors checked the validity of the generated themes. AA and KD 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.

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

Search strategy.

[\[DOCX File , 28 KB-Multimedia Appendix 1\]](#)

---

## Multimedia Appendix 2

Data extraction form.

[\[DOCX File , 19 KB-Multimedia Appendix 2\]](#)

---

## Multimedia Appendix 3

The metadata and population characteristics of each included study.

[\[DOCX File , 30 KB-Multimedia Appendix 3\]](#)

---

## Multimedia Appendix 4

Characteristics of the intervention in each included study.

[\[DOCX File , 29 KB-Multimedia Appendix 4\]](#)

---

## References

1. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. *Int J Epidemiol* 2014 Apr;43(2):476-493 [[FREE Full text](#)] [doi: [10.1093/ije/dyu038](https://doi.org/10.1093/ije/dyu038)] [Medline: [24648481](https://pubmed.ncbi.nlm.nih.gov/24648481/)]
2. Whiteford HA, Ferrari AJ, Degenhardt L, Feigin V, Vos T. The global burden of mental, neurological and substance use disorders: an analysis from the Global Burden of Disease Study 2010. *PLoS One* 2015;10(2):e0116820 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0116820](https://doi.org/10.1371/journal.pone.0116820)] [Medline: [25658103](https://pubmed.ncbi.nlm.nih.gov/25658103/)]

3. Jones SP, Patel V, Saxena S, Radcliffe N, Ali Al-Marri S, Darzi A. How Google's 'Ten Things We Know To Be True' could guide the development of mental health mobile apps. *Health Aff (Millwood)* 2014 Sep;33(9):1603-1611. [doi: [10.1377/hlthaff.2014.0380](https://doi.org/10.1377/hlthaff.2014.0380)] [Medline: [25201665](https://pubmed.ncbi.nlm.nih.gov/25201665/)]
4. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *Can J Psychiatry* 2019 Jul;64(7):456-464 [FREE Full text] [doi: [10.1177/0706743719828977](https://doi.org/10.1177/0706743719828977)] [Medline: [30897957](https://pubmed.ncbi.nlm.nih.gov/30897957/)]
5. Oladeji BD, Gureje O. Brain drain: a challenge to global mental health. *BJPsych Int* 2016 Aug;13(3):61-63 [FREE Full text] [doi: [10.1192/s2056474000001240](https://doi.org/10.1192/s2056474000001240)] [Medline: [29093905](https://pubmed.ncbi.nlm.nih.gov/29093905/)]
6. Murray C, Vos T, Lozano R, Naghavi M, Flaxman A, Michaud C, et al. 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. *Lancet* 2012 Dec 15;380(9859):2197-2223. [doi: [10.1016/S0140-6736\(12\)61689-4](https://doi.org/10.1016/S0140-6736(12)61689-4)] [Medline: [23245608](https://pubmed.ncbi.nlm.nih.gov/23245608/)]
7. Hester RD. Lack of access to mental health services contributing to the high suicide rates among veterans. *Int J Ment Health Syst* 2017;11:47 [FREE Full text] [doi: [10.1186/s13033-017-0154-2](https://doi.org/10.1186/s13033-017-0154-2)] [Medline: [28828036](https://pubmed.ncbi.nlm.nih.gov/28828036/)]
8. Blankers M, Schippers G. Internet therapy versus internet self-help versus no treatment for problematic alcohol use: a randomized controlled trial. *Suchttherapie* 2011 Aug 22;12(S 01). [doi: [10.1055/s-0031-1284519](https://doi.org/10.1055/s-0031-1284519)]
9. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. *BMC Psychiatry* 2010 Dec 30;10:113 [FREE Full text] [doi: [10.1186/1471-244X-10-113](https://doi.org/10.1186/1471-244X-10-113)] [Medline: [21192795](https://pubmed.ncbi.nlm.nih.gov/21192795/)]
10. Lucas GM, Rizzo A, Gratch J, Scherer S, Stratou G, Boberg J, et al. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Front. Robot. AI* 2017 Oct 12;4. [doi: [10.3389/frobt.2017.00051](https://doi.org/10.3389/frobt.2017.00051)]
11. Sebastian J, Richards D. Changing stigmatizing attitudes to mental health via education and contact with embodied conversational agents. *Comput Hum Behav* 2017 Aug 2;73(3):479-488. [doi: [10.1016/j.chb.2017.03.071](https://doi.org/10.1016/j.chb.2017.03.071)]
12. Hawton K, van Heeringen K. Suicide. *Lancet* 2009 Apr;373(9672):1372-1381 [FREE Full text] [doi: [10.1016/s0140-6736\(09\)60372-x](https://doi.org/10.1016/s0140-6736(09)60372-x)]
13. Anthes E. Mental health: there's an app for that. *Nature* 2016 Apr 7;532(7597):20-23. [doi: [10.1038/532020a](https://doi.org/10.1038/532020a)] [Medline: [27078548](https://pubmed.ncbi.nlm.nih.gov/27078548/)]
14. Abd-Alrazaq AA, Bewick B, Farragher T, Gardner P. Factors that affect the use of electronic personal health records among patients: a systematic review. *Int J Med Inform* 2019 Jun;126:164-175. [doi: [10.1016/j.ijmedinf.2019.03.014](https://doi.org/10.1016/j.ijmedinf.2019.03.014)] [Medline: [31029258](https://pubmed.ncbi.nlm.nih.gov/31029258/)]
15. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick B, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. *J Med Internet Res* 2020 Jul 13;22(7):e16021 [FREE Full text] [doi: [10.2196/16021](https://doi.org/10.2196/16021)] [Medline: [32673216](https://pubmed.ncbi.nlm.nih.gov/32673216/)]
16. Bendig E, Erb B, Schulze-Thuesing L, Baumeister H. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health – a scoping review. *Verhaltenstherapie* 2019 Aug 20:1-13. [doi: [10.1159/000501812](https://doi.org/10.1159/000501812)]
17. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. *J Med Internet Res* 2017 May 9;19(5):e151 [FREE Full text] [doi: [10.2196/jmir.6553](https://doi.org/10.2196/jmir.6553)] [Medline: [28487267](https://pubmed.ncbi.nlm.nih.gov/28487267/)]
18. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Method* 2005 Feb;8(1):19-32. [doi: [10.1080/1364557032000119616](https://doi.org/10.1080/1364557032000119616)]
19. Grant M, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J* 2009 Jun;26(2):91-108 [FREE Full text] [doi: [10.1111/j.1471-1842.2009.00848.x](https://doi.org/10.1111/j.1471-1842.2009.00848.x)] [Medline: [19490148](https://pubmed.ncbi.nlm.nih.gov/19490148/)]
20. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med* 2018 Sep 4;169(7):467. [doi: [10.7326/m18-0850](https://doi.org/10.7326/m18-0850)]
21. Search Tip. Google Scholars. 2020. URL: <https://scholar.google.com/intl/en/scholar/help.html> [accessed 2020-12-21]
22. Higgins J, Deeks J. *Selecting Studies and Collecting Data*. London, UK: John Wiley & Sons; 2008.
23. Orwin R, Vevea J. Evaluating coding decisions. In: *The Handbook of Research Synthesis and Meta-analysis*. Second Edition. New York, USA: Russell Sage Foundation; 2009:177-203.
24. Caton S, Chapman M. The use of social media and people with intellectual disability: a systematic review and thematic analysis. *J Intellect Dev Disable* 2016 Mar 9;41(2):125-139. [doi: [10.3109/13668250.2016.1153052](https://doi.org/10.3109/13668250.2016.1153052)]
25. Crampton NH, Reis S, Shachak A. Computers in the clinical encounter: a scoping review and thematic analysis. *J Am Med Inform Assoc* 2016 May;23(3):654-665. [doi: [10.1093/jamia/ocv178](https://doi.org/10.1093/jamia/ocv178)] [Medline: [26769911](https://pubmed.ncbi.nlm.nih.gov/26769911/)]
26. Darbyshire D, Baker P. A systematic review and thematic analysis of cinema in medical education. *Med Humanit* 2012 Jun;38(1):28-33. [doi: [10.1136/medhum-2011-010026](https://doi.org/10.1136/medhum-2011-010026)] [Medline: [22282424](https://pubmed.ncbi.nlm.nih.gov/22282424/)]
27. Freeman J, Caldwell P, Bennett P, Scott K. How adolescents search for and appraise online health information: a systematic review. *J Pediatr* 2018 Apr;195:244-55.e1. [doi: [10.1016/j.jpeds.2017.11.031](https://doi.org/10.1016/j.jpeds.2017.11.031)] [Medline: [29398062](https://pubmed.ncbi.nlm.nih.gov/29398062/)]
28. Smith V, Begley CM, Clarke M, Devane D. Professionals' views of fetal monitoring during labour: a systematic review and thematic analysis. *BMC Pregnancy Childbirth* 2012 Dec 27;12:166 [FREE Full text] [doi: [10.1186/1471-2393-12-166](https://doi.org/10.1186/1471-2393-12-166)] [Medline: [23270400](https://pubmed.ncbi.nlm.nih.gov/23270400/)]
29. Boyatzis R. *Transforming qualitative information: thematic analysis and code development*. London, UK: Sage Publications; 1998:302.

30. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006 Jan;3(2):77-101 [FREE Full text] [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)]
31. Howitt D. *Introduction to Qualitative Research Methods in Psychology*. London, UK: Pearson; 2016.
32. Wilson S, MacLean R. *Research Methods and Data Analysis for Psychology*. New York, USA: McGraw-Hill Higher Education; 2011.
33. Ali M, Rasazi Z, Mamun A, Langevin R, Rawassizadeh R, Schubert L. A virtual conversational agent for teens with autism: experimental results and design lessons. *arXiv* 2018:- epub ahead of print. [doi: [10.1145/3383652.3423900](https://doi.org/10.1145/3383652.3423900)]
34. Bresó A, Martínez-Miranda J, Botella C, Baños RM, García-Gómez JM. Usability and acceptability assessment of an empathic virtual agent to prevent major depression. *Expert Syst* 2016 May 25;33(4):297-312. [doi: [10.1111/exsy.12151](https://doi.org/10.1111/exsy.12151)] [Medline: [2016](https://pubmed.ncbi.nlm.nih.gov/2016/)]
35. Burton C, Szentagotai Tatar A, McKinstry B, Matheson C, Matu S, Moldovan R, Help4Mood Consortium. Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. *J Telemed Telecare* 2016 Sep;22(6):348-355. [doi: [10.1177/1357633X15609793](https://doi.org/10.1177/1357633X15609793)] [Medline: [26453910](https://pubmed.ncbi.nlm.nih.gov/26453910/)]
36. Cameron G, Cameron D, Megaw G, Bond R, Mulvenna M, O'Neill S. Assessing the Usability of a Chatbot for Mental Health Care. In: *International Conference on Internet Science*. 2019 Presented at: CIS'19; October 24-26, 2018; St. Petersburg, Russia p. 121-132. [doi: [10.1007/978-3-030-17705-8\\_11](https://doi.org/10.1007/978-3-030-17705-8_11)]
37. Fitzpatrick KK, Darcy A, Vierhile M. 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* 2017 Jun 6;4(2):e19 [FREE Full text] [doi: [10.2196/mental.7785](https://doi.org/10.2196/mental.7785)] [Medline: [28588005](https://pubmed.ncbi.nlm.nih.gov/28588005/)]
38. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (tess) to relieve symptoms of depression and anxiety: randomized controlled trial. *JMIR Ment Health* 2018 Dec 13;5(4):e64 [FREE Full text] [doi: [10.2196/mental.9782](https://doi.org/10.2196/mental.9782)] [Medline: [30545815](https://pubmed.ncbi.nlm.nih.gov/30545815/)]
39. Grolleman J, van DB, Nijholt A, van EA. Break the Habit! Designing an e-Therapy Intervention Using a Virtual Coach in Aid of Smoking Cessation. In: *International Conference on Persuasive Technology*. 2006 Presented at: CPT'06; May 18-19, 2006; Eindhoven, The Netherlands p. 133-141. [doi: [10.1007/11755494\\_19](https://doi.org/10.1007/11755494_19)]
40. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR Mhealth Uhealth* 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: [10.2196/12106](https://doi.org/10.2196/12106)] [Medline: [30470676](https://pubmed.ncbi.nlm.nih.gov/30470676/)]
41. Ku J, Han K, Lee HR, Jang HJ, Kim KU, Park SH, et al. VR-based conversation training program for patients with schizophrenia: a preliminary clinical trial. *Cyberpsychol Behav* 2007 Aug;10(4):567-574. [doi: [10.1089/cpb.2007.9989](https://doi.org/10.1089/cpb.2007.9989)] [Medline: [17711366](https://pubmed.ncbi.nlm.nih.gov/17711366/)]
42. Ly K, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. *Internet Interv* 2017 Dec;10:39-46 [FREE Full text] [doi: [10.1016/j.invent.2017.10.002](https://doi.org/10.1016/j.invent.2017.10.002)] [Medline: [30135751](https://pubmed.ncbi.nlm.nih.gov/30135751/)]
43. Pinto MD, Greenblatt AM, Hickman T, Rice HM, Thomas TL, Clochesy JM. Assessing the critical parameters of esmart-MH: a promising avatar-based digital therapeutic intervention to reduce depressive symptoms. *Perspect Psychiatr Care* 2016 Jul;52(3):157-168. [doi: [10.1111/ppc.12112](https://doi.org/10.1111/ppc.12112)] [Medline: [25800698](https://pubmed.ncbi.nlm.nih.gov/25800698/)]
44. Pontier M, Siddiqui G. A Virtual Therapist That Responds Empathically to Your Answers. In: *International Workshop on Intelligent Virtual Agents*. 2008 Presented at: WIVA'08; September 1-3, 2008; Tokyo, Japan p. 214-225. [doi: [10.1007/978-3-540-85483-8](https://doi.org/10.1007/978-3-540-85483-8)]
45. Razavi S, Ali M, Smith T, Schubert L, Hoque M. The LISSA Virtual Human and ASD Teens: An Overview of Initial Experiment. In: *International Conference on Intelligent Virtual Agents*. The LISSA virtual human and ASD teens; 2016 Presented at: CIVA'16; September 20-23, 2016; Los Angeles, USA. [doi: [10.1007/978-3-319-47665-0\\_55](https://doi.org/10.1007/978-3-319-47665-0_55)]
46. Schroeder J, Wilkes C, Rowan K, Toledo A, Paradiso A, Czerwinski M. Pocket Skills: A Conversational Mobile Web App To Support Dialectical Behavioral Therapy. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018 Presented at: CHI'18; April 12-14, 2018; Montreal, Canada. [doi: [10.1145/3173574.3173972](https://doi.org/10.1145/3173574.3173972)]
47. Poole-Wilson PA, Langer GA. Effect of pH on ionic exchange and function in rat and rabbit myocardium. *Am J Physiol* 1975 Sep;229(3):570-581. [Medline: [2014](https://pubmed.ncbi.nlm.nih.gov/2014/)]
48. Smith MJ, Ginger EJ, Wright K, Wright MA, Taylor JL, Humm LB, et al. Virtual reality job interview training in adults with autism spectrum disorder. *J Autism Dev Disord* 2014 Oct;44(10):2450-2463 [FREE Full text] [doi: [10.1007/s10803-014-2113-y](https://doi.org/10.1007/s10803-014-2113-y)] [Medline: [24803366](https://pubmed.ncbi.nlm.nih.gov/24803366/)]
49. Smith MJ, Humm LB, Fleming MF, Jordan N, Wright MA, Ginger EJ, et al. Virtual reality job interview training for veterans with posttraumatic stress disorder. *J Vocat Rehabil* 2015;42(3):271-279 [FREE Full text] [doi: [10.3233/JVR-150748](https://doi.org/10.3233/JVR-150748)] [Medline: [27721645](https://pubmed.ncbi.nlm.nih.gov/27721645/)]
50. Tanaka H, Sakti S, Neubig G, Toda T, Negoro H, Iwasaka H. Automated Social Skills Trainer. In: *Proceedings of the 20th International Conference on Intelligent User Interfaces*. 2015 Presented at: CIUT'15; 29 March-01 April 2015; Atlanta Georgia, USA. [doi: [10.1145/2678025.2701368](https://doi.org/10.1145/2678025.2701368)]
51. Tielman ML, Neerinx MA, Bidarra R, Kybartas B, Brinkman W. A therapy system for post-traumatic stress disorder using a virtual agent and virtual storytelling to reconstruct traumatic memories. *J Med Syst* 2017 Aug;41(8):125 [FREE Full text] [doi: [10.1007/s10916-017-0771-y](https://doi.org/10.1007/s10916-017-0771-y)] [Medline: [28699083](https://pubmed.ncbi.nlm.nih.gov/28699083/)]

52. Tielman ML, Neerinx MA, van Meggelen M, Franken I, Brinkman W. How should a virtual agent present psychoeducation? Influence of verbal and textual presentation on adherence. *Technol Health Care* 2017 Dec 4;25(6):1081-1096 [FREE Full text] [doi: [10.3233/THC-170899](https://doi.org/10.3233/THC-170899)] [Medline: [28800346](https://pubmed.ncbi.nlm.nih.gov/28800346/)]
53. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a relational agent by hospital patients with depressive symptoms. *Interact Comput* 2010 Jul 1;22(4):289-298 [FREE Full text] [doi: [10.1016/j.intcom.2009.12.001](https://doi.org/10.1016/j.intcom.2009.12.001)] [Medline: [20628581](https://pubmed.ncbi.nlm.nih.gov/20628581/)]
54. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a relational agent by hospital patients with depressive symptoms. *Interact Comput* 2010 Jul 1;22(4):289-298 [FREE Full text] [doi: [10.1016/j.intcom.2009.12.001](https://doi.org/10.1016/j.intcom.2009.12.001)] [Medline: [20628581](https://pubmed.ncbi.nlm.nih.gov/20628581/)]
55. Demirci H. User experience over time with conversational agents case study of woebot on supporting subjective well-being. Ankara, Turkey: Middle East Technical University; 2018.
56. DeVault D, Artstein R, Benn G, Dey T, Fast E, Gainer A. Sensesi Kiosk: a Virtual Human Interviewer for Healthcare Decision Support. In: Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems. 2014 Presented at: AAMS'14; 3-5 May 2014; Paris, France.
57. Elmasri D, Maeder A. A Conversational Agent for an Online Mental Health Intervention. In: International Conference on Brain Informatics. 2016 Presented at: CBI'16; 13-16 October 2016; Omaha, USA. [doi: [10.1007/978-3-319-47103-7\\_24](https://doi.org/10.1007/978-3-319-47103-7_24)]
58. Lisetti C, Amini R, Yasavur U, Rishe N. I can help you change! An empathic virtual agent delivers behavior change health interventions. *ACM Trans Manage Inf Syst* 2013 Dec 1;4(4):1-28. [doi: [10.1145/2544103](https://doi.org/10.1145/2544103)]
59. Milne M, Luerssen M, Lewis T, Leibbrandt R, Powers D. Development of a Virtual Agent Based Social Tutor for Children With Autism Spectrum Disorders. The 2010 International Joint Conference on Neural Networks (IJCNN). In: The 2010 International Joint Conference on Neural Networks. 2010 Presented at: IJCNN'10; July 2-7, 2010; Barcelona, Spain p. 18-23. [doi: [10.1109/IJCNN.2010.5596584](https://doi.org/10.1109/IJCNN.2010.5596584)]
60. Wargnier P, Benveniste S, Jouvelot P, Rigaud A. Usability assessment of interaction management support in LOUISE, an ECA-based user interface for elders with cognitive impairment. *TAD* 2018 Nov 26;30(3):105-126 [FREE Full text] [doi: [10.3233/tad-180189](https://doi.org/10.3233/tad-180189)]
61. Yasavur U, Lisetti C, Rishe N. Let's talk! speaking virtual counselor offers you a brief intervention. *J Multimodal User Interfaces* 2014 Sep 5;8(4):381-398. [doi: [10.1007/s12193-014-0169-9](https://doi.org/10.1007/s12193-014-0169-9)]
62. Luerssen M, Hawke T. Virtual Agents as a Service: Applications in Healthcare. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents. 2018 Presented at: CIVA'18; 20-23 November 2018; Sydney, Australia. [doi: [10.1145/3267851.3267858](https://doi.org/10.1145/3267851.3267858)]
63. Swartout W, Artstein R, Forbell E, Foutz S, Lane HC, Lange B, et al. Virtual Humans for Learning. *AIMag* 2013 Dec 15;34(4):13-30. [doi: [10.1609/aimag.v34i4.2487](https://doi.org/10.1609/aimag.v34i4.2487)]
64. Martínez-Miranda J, Bresó A, García-Gómez J. Look on the bright side: a model of cognitive change in virtual agents. In: International Conference on Intelligent Virtual Agents. 2014 Presented at: CIVA'14; 26-29 August 2014; Boston, MA, USA p. 285-294. [doi: [10.1007/978-3-319-09767-1\\_37](https://doi.org/10.1007/978-3-319-09767-1_37)]
65. Lahiri U, Bekele E, Dohrmann E, Warren Z, Sarkar N. Design of a virtual reality based adaptive response technology for children with autism. *IEEE Trans Neural Syst Rehabil Eng* 2013 Jan;21(1):55-64 [FREE Full text] [doi: [10.1109/TNSRE.2012.2218618](https://doi.org/10.1109/TNSRE.2012.2218618)] [Medline: [23033333](https://pubmed.ncbi.nlm.nih.gov/23033333/)]
66. Philip P, Micoulaud-Franchi J, Sagaspe P, Sevin ED, Olive J, Bioulac S, et al. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. *Sci Rep* 2017 Feb 16;7:42656 [FREE Full text] [doi: [10.1038/srep42656](https://doi.org/10.1038/srep42656)] [Medline: [28205601](https://pubmed.ncbi.nlm.nih.gov/28205601/)]
67. Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. *J Med Internet Res* 2018 Jun 26;20(6):e10148 [FREE Full text] [doi: [10.2196/10148](https://doi.org/10.2196/10148)] [Medline: [29945856](https://pubmed.ncbi.nlm.nih.gov/29945856/)]
68. Auriacombe M, Moriceau S, Serre F, Denis C, Micoulaud-Franchi JA, de Sevin E, et al. Development and validation of a virtual agent to screen tobacco and alcohol use disorders. *Drug Alcohol Depend* 2018 Dec 1;193:1-6. [doi: [10.1016/j.drugalcdep.2018.08.025](https://doi.org/10.1016/j.drugalcdep.2018.08.025)] [Medline: [30321739](https://pubmed.ncbi.nlm.nih.gov/30321739/)]
69. Yokotani K, Takagi G, Wakashima K. Advantages of virtual agents over clinical psychologists during comprehensive mental health interviews using a mixed methods design. *Comput Hum Behav* 2018 Aug;85(6):135-145. [doi: [10.1016/j.chb.2018.03.045](https://doi.org/10.1016/j.chb.2018.03.045)] [Medline: [2018](https://pubmed.ncbi.nlm.nih.gov/2018/)]
70. Fadhil A, Schiavo G. Designing for Health Chatbots. *ArXiv* 2019:1-10.
71. Laranjo L, Dunn A, Tong H, Kocaballi A, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc* 2018 Sep 1;25(9):1248-1258 [FREE Full text] [doi: [10.1093/jamia/ocy072](https://doi.org/10.1093/jamia/ocy072)] [Medline: [30010941](https://pubmed.ncbi.nlm.nih.gov/30010941/)]
72. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. *J Med Internet Res* 2019 Apr 5;21(4):e12887 [FREE Full text] [doi: [10.2196/12887](https://doi.org/10.2196/12887)] [Medline: [30950796](https://pubmed.ncbi.nlm.nih.gov/30950796/)]
73. Denecke K, Vaaheesan S, Arulnathan A. A mental health chatbot for regulating emotions (SERMO) - concept and usability test. *IEEE Trans Emerg Topics Comput* 2020:1-1. [doi: [10.1109/tetc.2020.2974478](https://doi.org/10.1109/tetc.2020.2974478)]
74. Albrecht U. Chances and Risks of Mobile Health Apps (CHARISMHA). 2016 May 26. URL: <https://nbn-resolving.org/urn:nbn:de:gbv:084-16051809293> [accessed 2020-12-21]

75. Singh K, Drouin K, Newmark LP, Lee J, Faxvaag A, Rozenblum R, et al. Many mobile health apps target high-need, high-cost populations, but gaps remain. *Health Aff (Millwood)* 2016 Dec 1;35(12):2310-2318. [doi: [10.1377/hlthaff.2016.0578](https://doi.org/10.1377/hlthaff.2016.0578)] [Medline: [27920321](https://pubmed.ncbi.nlm.nih.gov/27920321/)]
76. de Gennaro M, Krumhuber EG, Lucas G. Effectiveness of an empathic chatbot in combating adverse effects of social exclusion on mood. *Front Psychol* 2019;10:3061 [FREE Full text] [doi: [10.3389/fpsyg.2019.03061](https://doi.org/10.3389/fpsyg.2019.03061)] [Medline: [32038415](https://pubmed.ncbi.nlm.nih.gov/32038415/)]
77. Fitzpatrick M, Nedeljkovic M, Abbott J, Kyrios M, Moulding R. 'Blended' therapy: the development and pilot evaluation of an internet-facilitated cognitive behavioral intervention to supplement face-to-face therapy for hoarding disorder. *Internet Interv* 2018 Jun;12:16-25 [FREE Full text] [doi: [10.1016/j.invent.2018.02.006](https://doi.org/10.1016/j.invent.2018.02.006)] [Medline: [30135765](https://pubmed.ncbi.nlm.nih.gov/30135765/)]
78. Yu Z, Xu Z, Black AW, Rudnicki A. Chatbot Evaluation and Database Expansion via Crowdsourcing. In: Proceedings of the chatbot workshop of LREC. 2016 Presented at: LREC'16; 2016; UK.
79. Luxton DD. Ethical implications of conversational agents in global public health. *Bull World Health Organ* 2020 Jan 27;98(4):285-287. [doi: [10.2471/blt.19.237636](https://doi.org/10.2471/blt.19.237636)]
80. Pestian JP, Matykiewicz P, Linn-Gust M, South B, Uzun O, Wiebe J, et al. Sentiment analysis of suicide notes: a shared task. *Biomed Inform Insights* 2012 Jan 30;5(Suppl 1):3-16 [FREE Full text] [doi: [10.4137/bii.s9042](https://doi.org/10.4137/bii.s9042)] [Medline: [22419877](https://pubmed.ncbi.nlm.nih.gov/22419877/)]
81. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: systematic review. *J Med Internet Res* 2019 Nov 7;21(11):e15360 [FREE Full text] [doi: [10.2196/15360](https://doi.org/10.2196/15360)] [Medline: [31697237](https://pubmed.ncbi.nlm.nih.gov/31697237/)]
82. Kowatsch T, Otto L, Harperink S, Cotti A, Schlieter H. A design and evaluation framework for digital health interventions. *Inf Technol* 2019;61(5-6):253-263. [doi: [10.1515/itit-2019-0019](https://doi.org/10.1515/itit-2019-0019)]
83. Denecke K, Warren J. How to evaluate health applications with conversational user interface? *Stud Health Technol Inform* 2020 Jun 16;270:976-980. [doi: [10.3233/SHTI200307](https://doi.org/10.3233/SHTI200307)] [Medline: [32570527](https://pubmed.ncbi.nlm.nih.gov/32570527/)]
84. Abd-Alrazaq A, Safi Z, Alajlani M, Warren J, Househ M, Denecke K. Technical metrics used to evaluate health care chatbots: scoping review. *J Med Internet Res* 2020 Jun 5;22(6):e18301 [FREE Full text] [doi: [10.2196/18301](https://doi.org/10.2196/18301)] [Medline: [32442157](https://pubmed.ncbi.nlm.nih.gov/32442157/)]
85. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. *J Med Internet Res* 2014 Jun 5;16(6):e146 [FREE Full text] [doi: [10.2196/jmir.3077](https://doi.org/10.2196/jmir.3077)] [Medline: [24905070](https://pubmed.ncbi.nlm.nih.gov/24905070/)]
86. Kan S. *Metrics and Models in Software Quality Engineering*. Boston: USA: Addison-Wesley Longman Publishing Co., Inc; 2002.
87. Neary M, Schueller S. State of the field of mental health apps. *Cogn Behav Pract* 2018 Nov;25(4):531-537 [FREE Full text] [doi: [10.1016/j.cbpra.2018.01.002](https://doi.org/10.1016/j.cbpra.2018.01.002)] [Medline: [33100810](https://pubmed.ncbi.nlm.nih.gov/33100810/)]

## Abbreviations

**AI:** artificial intelligence

**mHealth:** mobile health

**PRISMA:** Preferred Reporting Items for Systematic reviews and Meta-Analyses

*Edited by G Eysenbach; submitted 16.01.20; peer-reviewed by M Leary, M Lahti, YC Wang, E Broglia, J Apolinário-Hagen; comments to author 24.03.20; revised version received 01.06.20; accepted 21.06.20; published 13.01.21*

*Please cite as:*

Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M

*Perceptions and Opinions of Patients About Mental Health Chatbots: Scoping Review*

*J Med Internet Res* 2021;23(1):e17828

URL: <http://www.jmir.org/2021/1/e17828/>

doi: [10.2196/17828](https://doi.org/10.2196/17828)

PMID: [33439133](https://pubmed.ncbi.nlm.nih.gov/33439133/)

©Alaa A Abd-Alrazaq, Mohannad Alajlani, Nashva Ali, Kerstin Denecke, Bridgette M Bewick, Mowafa Househ. Originally published in the *Journal of Medical Internet Research* (<http://www.jmir.org>), 13.01.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the *Journal of Medical Internet Research*, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.