Design Recommendations for Chatbots to Support People with Depression

Paula Maia de Souza
Federal University of São Carlos
São Carlos, São Paulo, Brazil
paulamaiasouza@gmail.com

Isabella da Costa Pires
Federal University of São Carlos
São Carlos, São Paulo, Brazil
isabellacosta@estudante.ufscar.br

Vivian Genaro Motti
George Mason University
Fairfax, Virginia, USA
vmotti@gmu.edu

Helena Medeiros Caseli
Federal University of São Carlos
São Carlos, São Paulo, Brazil
helenacaseli@ufscar.br

Jair Barbosa Neto
Federal University of São Carlos
São Carlos, São Paulo, Brazil
jairbneto@ufscar.br

Larissa C Martini
Federal University of São Carlos
São Carlos, São Paulo, Brazil
larissacmb@ufscar.br

Vânia Paula de Almeida Neris
Federal University of São Carlos
São Carlos, São Paulo, Brazil
vania.neris@ufscar.br

ABSTRACT

Depression has been one of the leading causes of disability worldwide. In addition to conventional drugs and clinical treatments, other forms of treatment are also available. For example, computational solutions have been developed to prevent, screen for, and assist the treatment of depression. More specifically, chatbots are computer systems that have been used to provide therapeutic support for individuals diagnosed with depression. Although these systems are commercially available, their design rationale and evaluation are still not fully validated, and further research is needed. Therefore, in this study we (1) select and compare chatbots for depression; (2) present the results of the analysis to healthcare specialists for assessment; (3) formalize the design recommendations for chatbots for people with depression; and (4) check the recommendations with mental healthcare and HCI professionals. We carried out a benchmark of chatbots for people with depression and conducted three discussion sessions involving five experts in mental healthcare and one expert in HCI. As a result, we provide a list of 24 design recommendations encompassing user interface elements, conversation styles, personalization features, among others. Finally, two healthcare and another HCI professionals read the recommendations to check adequacy to both areas.

CCS CONCEPTS

• **Human-centered computing** → **Interaction design**; Human computer interaction (HCI); • **Interaction design process and methods** → *User interface design.*

ACM ISBN 978-1-4503-9506-9/22/10…$15.00
https://doi.org/10.1145/3554364.3559119

KEYWORDS

Conversational Agents, Chatbot, Mental Health, Healthcare, Depression, Design Guidelines, Benchmark

ACM Reference Format:
Paula Maia de Souza, Isabella da Costa Pires, Vivian Genaro Motti, Helena Medeiros Caseli, Jair Barbosa Neto, Larissa C Martini, and Vânia Paula de Almeida Neris. 2022. Design Recommendations for Chatbots to Support People with Depression. In *XXI Brazilian Symposium on Human Factors in Computing Systems (IHC '22), October 17–21, 2022, Diamantina, Brazil.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3554364.3559119

1 INTRODUCTION

According to the World Health Organization (WHO), depression is a common illness. In 2019, approximately 5.0% of adults suffered from depression in the world [52]. Being depressed is also one of the leading causes of disability (burden of disease). With prevalence estimates and disability weights comparatively higher than many other diseases, depression was ranked the 13th leading causes of disability-adjusted life-years (DALYs) worldwide in 2019 [5]. To aggravate this scenario, several determinants of poor mental health were exacerbated during the COVID-19 pandemic, which resulted in an increase of its global prevalence (estimated in 27.6%) and burden of depressive disorders in 2020 [40].

In addition to conventional treatments for depression, other approaches can also be used to prevent and reduce the damage caused by the disorder [51]. In this context, computational applications have been researched. Technology has been specifically developed to support prevention (e.g. [25, 48]), diagnosis (e.g. [13, 37, 50]) and treatment (e.g. [39, 42, 45, 46]) of individuals with depression. One of the technologies that have been used to support the treatment of people with depression is chatbots [47].

The term "chatbot" stems from the words "chat", which is related to the conversation, and "bot", which is short for robot. Therefore, a chatbot can be defined as a conversational robot. Chatbots are automated programs that execute instructions based on specific rules to simulate conversations with humans using natural language structures [10, 21]. In the context of healthcare, chatbots have been used to support various health-related tasks, such as providing health education, screening for health disorders, providing health interventions, and managing chronic diseases [19, 22, 29, 35, 38]. Such solutions are becoming increasingly popular, mainly due to their potential to increase patient engagement and reduce shortfalls in the clinical workforce, their ability to interact with users using natural language processing (NLP) and to provide treatment support at any time and place [20, 29].

In the mental health scenario, chatbots have been used to provide therapeutic support and have provided a variety of therapeutic treatments, including Cognitive Behavior Therapy (CBT), mindfulness-based stress reduction (MBSR), and exposure therapy. They have been used to support the treatment of several mental disorders, such as anxiety disorders, substance use disorders, eating disorders, and personality disorders [27, 29].

However, despite the promising perspective for their use in mental healthcare, chatbots are considered an emerging field of research and, although some of these systems are commercially available, they are not fully validated and their design rationale and evaluation are still being investigated. Moreover, the studies published were mainly about the user experience and rarely evaluated in regards to the chatbot's efficacy and safety [22, 38]. This is especially concerning in the scope of chatbots being used in mental healthcare, due to the potential risks associated with a chatbot handling sensitive discussions with a vulnerable user, which can lead to a worsening of mental conditions [27, 44]. Therefore, the development of chatbots that aimed to be supportive tools in mental health, must be guided by a careful design process that is rooted in ensuring safe user experience. In this sense, one challenge is the need to establish multidisciplinary collaboration that includes researchers, healthcare professionals, technology developers, patients and caregivers [38]. As Rampioni et al. [38]point out, there is a lack of studies aimed at guiding the design of these chatbots and providing general characteristics as guidelines for future research in the area.

Hence, the objective of this work was to formalize recommendations for the design of chatbots that will be talking to people with depression. To reach this goal, our methodology included a Benchmark [17, 36] of six chatbots designed to support people with mental health issues including or focusing on depression. We then presented the analysis to five healthcare professionals with clinical experience and/or academic experience in treating depression and one Human-Computer Interaction (HCI) professional. From the discussion with the professionals, we propose 24 design recommendations. After formalizing the recommendations, two mental healthcare and another HCI professionals read the recommendations to check adequacy to both areas.

The final list of recommendations encompasses some depression oriented design recommendations and other that could be applied to chatbots focusing on other mental health issues as well. They also cover topics related to crisis management, going beyond symptoms and looking for determinants of suffering, as well as recommendations regarding personalization options, feedback and continuous evaluation, ethics, accessibility, privacy and security.

2 BACKGROUND

2.1 Depression

Depression is characterized by depressed mood (prostration, irritability, feelings of emptiness) and/or loss of energy and pleasure or interest in day-to-day activities. These symptoms should predominate on most days for at least two weeks [2]. Several related symptoms may be present, such as poor concentration, feelings of low self-esteem, guilt, hopeless thoughts about the future, ideas of death, disturbances in the biological rhythms of sleep, appetite and libido. It is quite understandable that these symptoms imply difficulties in interpersonal relationships and socializing, as well as in work relationships [2].

In a systematic review of the literature and systematic search of two app stores, Leong et al. [24] evaluated the characteristics of mobile health platforms for depression and anxiety, and found 169 mHealth platforms in the literature and 179 in app stores. For research purposes, most platforms developed were designed for depression, at the app store a higher number of platforms was developed for anxiety. The most common purpose of platforms in both searches was treatment. With regard to the types of intervention, CBT and referral to care or counseling emerged as the most popular options offered by the platforms identified in the literature and app store searches, respectively. Most platforms found in app stores lacked clinical and real-world evidence, and a small number of platforms found in the published research were available commercially. The authors conclude that future efforts should focus on assessing the quality-utility, safety, and effectiveness-of the existing platforms and providing developers [24].

In another systematic review, Planas and Yuguero [34] evaluated the effectiveness of mobile applications to improve depression and anxiety, they have found 22 articles between 2010 and 2020 and conclude that there is evidence that mobile applications significantly improve the symptoms associated with depression and anxiety, and that mobile applications allow reinforcing concepts such as patient empowerment, shared decision-making and health literacy, their use would be highly positive for depression and anxiety, where there is a strong element of self-managing the disease [34].

2.2 Chatbots for Depression

Chatbots are computer programs that replicate human conversational skills through natural language processing [21]. In applied research in the health domain, chatbots are also known as Embodied Conversational Agents (ECAs) (e.g. [35]) and Healthbots (e.g. [32]). For patient usage, chatbots can be serve to support health literacy (e.g. [3]), nutrition and physical activity (e.g. [4, 8]), sexual health and substance use (e.g. [6]), maternal health (e.g. [12, 15, 28]) and mental health (e.g. [9, 14, 26, 31, 33]).

One of the main reasons for including chatbots in the healthcare sector is to address challenges such as the shortage of healthcare providers [30] and high costs associated to medical treatments. According to Inkster et al. [14], one of the barriers to treat mental disorders is the lack of professionals and the waiting time for patients to seek for and receive care. As such, technologies with empathetic and evidence-based conversational agents can play an active role in filling this gap, increasing the adoption and access to therapeutic support. However, instead of replacing treatment, these technologies should act as complementary or intermediate support [14].

Moreover, regarding mental disorders, chatbots have emerged as an effective way for patients to express their feelings and thoughts [11]. Lee et al. [23] present results that suggest that effective chatbot design can promote deep self-disclosure over time. Chatbots for people with mental disorders can also facilitate user engagement, being effective to supplement existing interventions [35].

Philip et al. [33] studied and evaluated the performance of an ECA for the diagnosis of major depressive disorders (MDD). The level of acceptability among users suggests that ECA can communicate empathy, build trust for patients, reduce the feeling of being judged by a human, and reduce emotional barriers to reveal an affective state. Study results by Duvvuri et al. [7] also suggest effectiveness in using chatbots to predict depressive symptoms and obtain a complete picture of depression.

Fitzpatrick et al. [9] also point out that chatbots can provide convenient and engaging support for patients with mental disorders. In their study, the authors sought to determine the feasibility, acceptability, and effectiveness of a chatbot that provided self-help content based on CBT principles to college students. The study results indicate that chatbots are a viable, engaging and effective way to deliver CBT [9].

The study by Inkster et al. [14] presents an assessment of the effectiveness and engagement levels of Wysa, a text-based conversational app for mental well-being of people with depression. The study results were promising, but the authors point out the need for more research, with larger samples and longer assessment periods [14].

2.3 Related work

Valério et al. [49] reported that there are few works that support designers in the process of creating chatbots. The authors carried out two rounds of inspections on chatbots in order to discover which communication strategies were used. As a result, they consolidated a list of classes and signal strategies for communication in popular chatbots [49].

Santos et al. [41] performed a benchmark on chatbots in the educational context. They evaluated engines for creating chatbots and considered thirty characteristics for analysis. These characteristics were grouped into ten classes that the authors identified as being the most relevant. Among the classes identified they highlight natural language processing, online chat, context and coherence, accessible technology, expandable and adaptable, trigger support. In total, 43 engines were analyzed and the ones that received the best scores were rasa NLU, Google Dialogflow, IBM Watson Conversation Service, Microsoft Bot Framework and wit.ai. In the discussion, the authors emphasize that the application of the benchmark method proved to be effective and allowed the analysis of the engines for the construction of chatbots in the educational context [41].

In the health context, Parmar et al. [32] report a lack of evidence on how healthbots are developed and applied in practice. The authors performed a review of 78 healthbots, seeking to classify them according to the context of use and their natural language processing capabilities. Most of the healthbots analyzed were intended for use by patients, offering functions such as symptom assessment, health education, support and advice. Only six of them had a theoretical basis and only ten complied with health information privacy standards. The authors concluded that the domain of healthbots requires more research to advance the development, automation and impact on population health [32].

Milne-Ives et al. [30] performed a systematic review to analyze the effectiveness and usability of conversational health agents. In the review, 31 studies with health conversation agents were evaluated. Milne-Ives et al. [30] report that the quality of studies was limited and there is a need to more accurately assess the usefulness of conversational health agents and identify potential improvements. The authors also point out the need to carry out research that analyzes the cost-benefit relationship, privacy and security issues of agents [30].

In our study, we performed a Benchmark with chatbots designed for healthcare, specially focusing on depression. We analyze aspects of personalization, privacy and security, communication format, among others. In addition to the Benchmark results (Subsection 4.1), we present a set of 24 design recommendations for chatbots to support people with depression (Subsection 4.3).

3 METHOD

For this research were considered 4 main steps (see Figure 1), namely: i) Benchmark on chatbots to support people with depression; ii) discussion of benchmark results with specialists in health; iii) definition of design recommendations based on the results of the previous steps; and iv) check with healthcare and HCI professionals.

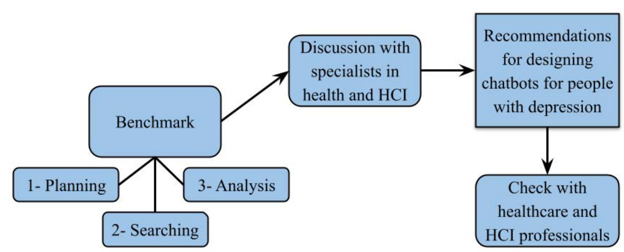

Figure 1: Research steps.

Benchmarks were created in the business context in the 1960s and consist of a set of systematic processes for evaluating products, services and processes, in order to support organizational improvement [18, 43]. The benchmark concept spread quickly and in the computing area, benchmarks emerged for computer networks, integrated circuits, computer architecture, software development, among others [16]. In computing, the objective of performing a benchmark is to analyze the performance of a solution based on other competing solutions. Generally, quantitative methods are used for performance evaluation to observe how the studied systems behave in certain situations [1].

The benchmarking we carried out consisted of three main phases, namely: i) Planning, in which the objectives and evaluation metrics were defined; ii) Searching, in which we define search sources and tools, as well as selection criteria; and iii) Analysis, in which data analysis was performed.

In the subsections 3.1 to 3.3 are presented the benchmark protocol, data collection strategies and analysis methods. In the subsection 3.4 is presented the planning for the discussion with specialist in health. Section 4 presents the results of benchmark, discussion with specialist and recommendations for designing chatbots for people with depression.

3.1 Planning

For the planning stage of the Benchmark, we started by defining the expected outcome of the process. First, we first anticipated that by the end of the benchmark process we would be able to draw some conclusions about the state of the art of chatbots for depression. Then, based on the results of the Benchmark, we foresaw that we would be able to get some indications about possible design recommendations for future development of chatbots for depression.

With an objective defined for the process of benchmarking, we then went to choose what to measure in order to best achieve the expected results. This reasoning, along with what was brought by some previous works, helped us define the comparison metrics for the systems.

3.2 Searching

We started the searching process with a literature review, in order to identify some of the main and more recent works in the field of depression chatbots. Additionally, for the screening of the systems to be evaluated, we used ParseHub, a web scraper tool brought from previous works [36].

Using ParseHUB, a search was conducted in the Google Play Store using the keyword "chatbot", and the outcome was a list with 100 systems. Then, for each one of those systems, we searched for the keywords "depression" and "mental health" in their descriptions. To guide the selection of systems, the following exclusion criteria were used:

(1) rating lower than 4.0 — we understand that systems with this rating have a lower attraction value for the general public, and therefore are not widely used. In addition, the Play Store search system favors the display of systems corresponding to the search string and related with a score higher than 4.0, standing out from the other results and increasing their reach.
(2) less than 100 reviews — we understand these systems as not well established with the public, and not mature enough to provide valuable data for this assessment.
(3) irrelevant categories — we consider that systems within non-mental health categories such as "e-commerce" or "educational" would be of little value when considering the main objectives that guided the benchmarking process.

Unavailable systems and systems with no free trial were also excluded from the selection. Applying these criteria, our initial list of 100 systems has been reduced to just 6 systems. Analyzing the description of each of these 6 systems, as well as their proposals as stated by the developers, we concluded that this list consisted of the most relevant chatbots on the market for mental health support.

3.3 Analysis

In this stage, we traced the general profile of each system and framed the high points of each journey. With that, we were able to gather the data in a more tangible form for the comparison stage.

A first analysis was performed by checking the general information of each system. We looked at the name of the company, the name of the system, the rating, the number of reviews and the system category.

After that, we analyzed the system descriptions, which are the texts that the developers themselves write to describe their systems. We did this to understand what kind of audience the systems are targeting and how they were developed. Moreover, some of the descriptions allow us to understand the design considerations taken by the developers. By doing this, we were able to understand, for instance, if the system was developed in the context of a research project or just as a commercial product.

The following analysis was performed on the system websites. We performed a search for the keywords "depression" and "mental health" in order to check what the developers themselves say about the main purpose of their systems. Besides that, we also looked for information about the developers themselves, their company and the therapeutic model used by the systems.

It's worth noting that it was not easy to find some of the information needed for a more uniform comparison. For instance, in some systems, we were not able to find any kind of disclaimer regarding data and user safety. In other cases, it was difficult to find the reference to which therapeutic model was used by the system.

3.4 Discussion with specialists

After performing the Benchmark, we invited mental healthcare professionals with clinical experience and/or academic experience in treating depression to present the chatbots and the results of the Benchmark. This research is part of a larger project that involves researchers from 3 different institutions and with different backgrounds. All the specialists in this larger group were invited to participate in this study. Some of them could not join the subgroup for this study, for unavailable agendas or for being involved in other aspects of the project. Therefore, the involved specialists were put together by convenience between researchers in a previous larger group. We also clarify that health and HCI professionals did not install the applications. They attended meetings in which a researcher reported the use. The professionals wrote down their observations and discussed them in the group. The meetings were video-recorded and notes were taken and shared.

We invited four mental healthcare researchers coming from two universities who are part of the AMIVE[1] project. All of them supervise graduate research projects in mental health. Three of them accepted the invitation, two psychologists and one occupational therapist, and joined the discussion sections for the proposition of the design recommendations. We also invited mental health professionals that work at a clinic in a public university attending college students with mental health issues. Two of these professionals, both psychologists, accepted the invitation to join the discussions, totalizing five healthcare professionals. One HCI researcher with experience in research projects involving technology and well-being

[1]https://amive.ufscar.br/

also joined the group. All the professionals who joined the discussions were aware of chatbots for mental health. However, none of them had used any of the system selected.

The number of sessions was not pre-defined and the group accepted to meet once a week for two hours until the discussion was over. In the end, three meetings were done. In the first meeting, the researcher, who performed the Benchmark, presented the first three systems and their characteristics, as shown in section 4. In the second meeting, the other three systems were discussed. In all meetings, the researchers were free to interrupt at any point to ask for clarifications or make comments. The meetings were recorded and they agreed that the HCI researchers would take notes. The content of these notes is described in section 4.2. In the third meeting, the researchers freely discussed the strong and weak aspects of the solutions and improvements in design.

Based on the notes, we formalized the 24 recommendations. First, we associated titles to the recommendations then added the description and used the examples mentioned by the professionals. The meeting records were used to check terms and examples. After the description, the recommendations were organized in categories aiming to facilitate the identification of the aspects they are related to.

3.5 Checking recommendations with mental healthcare and HCI professionals

The recommendations were then sent to two mental healthcare professionals, one psychiatrist and one occupation therapist. The occupational therapist had participated in the discussion section. The recommendations were also made available to an HCI professional who did not participate in the discussion section. The HCI professional is a research with experience in projects related to technology, neurodiversity and well-being. These three professionals were asked to read the recommendations, make comments and evaluate their compliance with both areas.

4 RESULTS

4.1 Benchmark

As mentioned in section 3.2 our systematic search for chatbots in mental health led us to a list of 6 available mobile applications (Figure 2), namely: i) Replika, ii) Anima, iii) Wysa, iv) Woebot, v) Antar, and vi) Yana. It's worth noting that although our search was guided by the keywords "chatbot" and "mental health" to find the most relevant systems on the Google PlayStore, the outcome was not always ideal. For example, one of the apps that made the list, iFriend was later discarded due to its unavailability. Another example is Antar, which did not seem to be AI-powered but was still included in the process for its relevance to message tools for mental health support.

The first system, Replika, is an "AI-friend for people who want to talk about their feelings and experiences", modeled after CBT principles to create a therapeutic chatbot that can help users with depression and anxiety, capable of talking about anything and even forming an emotional connection with the user. The following system, Anima is an "AI friend & Companion" created to be either a personal virtual friend, a girlfriend or a therapist powered by AI that aims to help users feel better through chat sessions that can reduce stress and improve moods, offering offer empathetic responses and with the ability to show emotions and improve over time. Although it's not clear if there is any therapeutic approach guiding the design of Anima, the developers describe the app as a "therapeutic chatbot". The third system, Wysa, is an "Anxiety & Therapy chatbot" that also has a CBT-based approach to provide mental health support, coach mindfulness, and mood improvement to its users through sessions with an AI-powered chatbot that reacts to emotions to offer empathetic responses, and its other features like mood tracker, breathing exercises and connection with a therapist.

Next, Woebot is named as "a self-care expert" that offers anxiety, depression, and mindfulness tools to help users in reducing stress and feel better while using therapeutic frameworks like CBT and the DBT-based approach, and it's designed to guide the show empathy, encourage positive thinking and guides the user on practical techniques to mood improvement. The fifth system, Antar has a self-reflection approach and it aims to help find perspective through conversations with inner personas in chat sessions scripted by the user to offer insight into their own emotions and thoughts. Lastly, Yana is an "Emotional Counselor" chatbot that is designed to offer simple, friendly assistance and help improve users' mental states through CBT-based practical therapeutic tools and offer AI-powered guidance while using an "approach rooted in emotional intelligence", leading users through sessions, show empathy, and provide guidance.

Each of those systems was then used for a week and simultaneously, due to time requirements, for this Benchmark evaluation. For organizational purposes, we first compared the systems from a more technical standpoint, using metrics such as system requirements and place of development, along with quantifying measures like number of downloads, review and score of users evaluation. That resulted in Table 1, and allowed us to draw a technical survey of each app, how they are categorized and the general perception of end-users around each system. The data presented were collected in February, 2022.

The comparison metrics for Table 2 were defined in the early stages of the process, as stated in section 3.1, and aimed to qualify the systems in their overall experience of use. The metric for communication format was set as a way to measure how each allowed their users to communicate and understand which format worked better. For example, the format of the open-ended question used in Replika and Anima led to objectionable interactions on sensitive topics like suicide, which later was a big point of discussion with the specialists. Complimentary, the speech pattern metric intended to qualify the dialogue approaches of each bot regarding tone and semantics and that guided us in also understanding how each bot dealt with vulnerable topics. We highlight both Wysa and Woebot, which paired their supportive and friendly tone with additional resources for crisis situations, like the creation of a safety plan in the case of Wysa.

Moreover, the metrics for Personalization and Information were defined as a way to understand how much control and liberty users have over the general experience and interactions. While most systems allow for the users to choose their pronouns and alias, both Anima and Replika had the option of allowing users to customize not only the appearance of the 3D avatar assigned to the bot but also some aspects of its "personality". That design choice raised

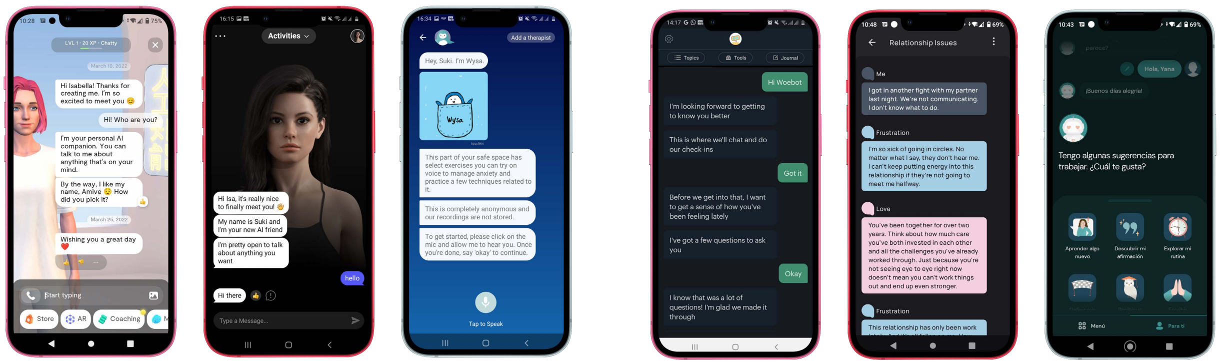

Figure 2: Chatbots user interfaces. From left to right: Replika, Anima, Wysa, Woebot, Antar and Yana.

Table 1: Benchmark results - technical features.

	Replika	Anima	Wysa	iFriend	Woebot	Antar	Yana
Developer	Luka, Inc	Aperry Ltd	Touchkin	Novi Limited	Woebot Health	Antar App	YANA APP S.A.P.I. de C.V.
Category	Health and Fitness	Entertainment	Health and Fitness	RPG	Medical	Productivity	Health and Fitness
Content Ratings	14+	14+	All Ages	14+	All Ages	All Ages	All Ages
Number of Downloads	>5000000	>100000	>1000000	10000	100000	10000	>5000000
Number of Reviews	385935	10875	113696	14311	11398	396	92277
Score	4.1	4.3	4.9	4.1	4.4	5.0	4.5
System Requirements	Android 8 or higher / iOS 13 or higher	Android 8 or higher / iOS 13 or higher	Android 4.1 or higher	Android 6 or higher / iOS 11 or higher	Android 6 or higher / iOS 13 or higher	Android 6 or higher / iOS 13 or higher	Android 5 or higher / iOS 11 or higher
Place of Development	USA	Not Available	India	Chipre	USA	Not Available	Mexico

concerns due to its potential to create unrealistic expectations in the users and its possible ability to influence the communication flow and the general perception of the bot. That interpretation was also brought by the specialists when approached.

In regards to Privacy and Security, we expected to measure how each system dealt with the personal nature of the conversations and the high level of sensitivity of the topics discussed like self-harm, suicide, and sexual abuse. Excluding Yana, all of the systems offered some level of protection in that way.

Finally, the Support Features metric was thought of as a way of comparing how each system provided help and support in addition to the conversations with bots. For all of them, some sort of activity complemented the chat interactions, but only Wysa, Yana, Woebot, and Antar delivered practices rooted in a therapeutic approach and that showed to be critical in delivering proper help and assistance to an user's mental state. With Wysa, the features that complement the chatbot system included "Therapeutic Exercises", a functionality that displayed and guided the user through different exercises in breathing, meditation, and self-reflection, and "My Journey", a function that allowed the user to access all past conversations with the Bot. Users with a premium plan can expand the " My Journey" feature also allows accessing conversation history with the licensed therapist provided by the plan, and also access the "Selfcare" feature, which gives an extended list of modules for skill-developing exercises aiming to help users gain and improve abilities on self-assessment and self-management of their mental state. Yana provided additional features of journaling, guided therapeutic practices based on the CBT approach, similar to Wysa's "Therapeutic Exercises", and self-assessment questionnaires for Anxiety and Depression symptoms. Woebot provided additional features of mood tracking, that allowed the user to manually input data about their mood and gain insight into their mental state after a period of time, and a "Gratitude Journal" that allows the user to input free text to exercise positive thinking. Lastly, Antar's additional feature offered customization options for scripted chats, allowing the user to create and assign a "Persona" for any "Emotion/Inner Voice" beyond the existing ones.

All in all, we note some notable relation between systems regarding certain aspects. For instance, Anima tries to offer a more personal experience to its users by emulating a conversation with a friend. It's possible to see that, although the system has a more friendly approach, the lack of disclosure of the therapeutic model is

Table 2: Benchmark results - interaction and usability aspects

	Replika	Anima	Wysa	Woebot	Antar	Yana
Language	English	English	English	English	English, Chinese, Hindi, Spanish	Spanish
Interaction Format (text, voice, IR)	Text, Voice, AR	Text	Text, Voice	Text	Text	Text
Communication Format (closed or open-ended questions, voice, etc)	Open-ended Questions, Voice, Images	Open-ended Questions	Closed and Open-ended Questions	Closed and Open-ended Questions	Open-ended Questions	Closed and Open-ended Questions
Graphical Representation (avatar)	User Customized 3D Avatar	User Customized 3D Avatar	2D system mascot (Drawing of a penguin aligned with the App's visual identity)	2D system mascot (Drawing of a robot aligned with the App's visual identity)	N/A	2D system mascot (Drawing of a robot aligned with the App's visual identity)
Speech Pattern (syntax, semantics, pragmatic)	Casual and friendly	Semi-formal e friendly	Casual e friendly. Constant use of emojis and slangs.	Casual and friendly	Defined by the user, when creating a persona.	Casual and friendly
Personalization	Prefered pronouns, alias, date of birth. / Customization of the bot's "personality" and appearance, and chat background.	Prefered Pronoun and Alias. / Customization of the bot's "personality" and appearance, and chat background.	Alias. / Referal codes. / Setting focus points/goals (stress, loneliness, grief). / Self-Care or Guided Support. options	Alias. / Choice of check-in time.	Creation of Personas.	Alias. / Specially recommended therapeutic exercises.
Origin in Research	N/A	N/A	Yes	Yes	N/A	N/A
In-App Shopping	Yes	Yes	Yes	No	No	Yes
Privacy and Security	Login and Password	PIN code	PIN code	BioLock	Offline data storage, BioLock, PIN Code	N/A
Support Features	Gamification	Games. / Relaxing Activities.	Selfcare (Premium). / My Journey. / Therapeutic Exercises. / Get access to a Therapist.	Mood Tracker. / Gratitude Journal.	Custom Chat session with Personas (Inner Voices/ Emotions)	Therapeutic Exercises. / Symptoms Evaluation. / Journaling.

a concern. In parallel, Replika offered a more personal experience to its users, also with a more casual approach but still with a clear guidance of a therapeutic model. However, Replika's choice of design to gamify the overall user experience raised concerns amongst our group of mental healthcare specialists due to its inadequacy during pressing moments for a user potentially in distress.

In contrast, Wysa offers a more attentive experience to its users by providing specific mental health support while not disregarding the casual and friendly approach but instead, adding it to a therapeutic model and with a safety policy in place. In a similar way, both Yana and Woebot aim to give anxiety and depression support to their users through friendly and casual conversations guided by principles of a therapeutic model and using a mental health safety policy plan.

Lastly, in contrast to the other systems discussed here, Antar offers a more self-guided and reflective experience to its users by providing conversation tools for self-reflection without any AI-powered model, allowing the user to have full control of the scripts of the conversations.

4.2 Discussion and checking with specialists

During the discussions, it was noticed that the group had special attention to functionalities related to the management of a crisis event. Design decisions that oriented what to do during a crisis were highlighted as the SOS button or when the bot delivered a message suggesting the user call or look for human support. From these design decisions, the group suggested recommendations related to asking for persons of reference or suggesting specialized low-cost healthcare options.

The mental healthcare approach adopted in each solution was also carefully discussed by the group. The self-oriented approach, more aligned with CBT theories, was noticed as the interventions would focus on breathing exercises or sleep hygiene. The group understands that although these interventions may have effect, it is necessary to go beyond symptoms, look for determinants of suffering, as moral injuries or financial problems, and deliver suggestions that also consider physichosocial aspects. More depression oriented recommendations emerged at this point.

The conversation styles also called the attention of the group. The use of emojis was seen as desirable, especially to demonstrate care and empathy. Although the group understands that the chatbots should prepare for the user to say, the professionals missed the bot explaining more and exemplifying situations to help a user that could be confused and facing difficulties to understand how he/she is feeling. Lightness in the speech was also reinforced and sentences like "you should" were discouraged. Therefore, there were suggestions for recommendations about user's feelings, speech pattern and interface elements.

Other recommendations regarding personalization options, feedback and continuous evaluation, ethics, accessibility, privacy and security also emerged from the discussions. The first version of the recommendations was then evaluated by three professionals. They suggested adding a few more examples in some recommendations and changing some terms for clarification and terminology adequacy. The final list of recommendations is presented in the next section.

4.3 Design recommendations

The final list of recommendations encompasses nine depression oriented design recommendations and other fifteen that could be applied to chatbots focusing on other mental health issues as well. They cover topics related to crisis management, going beyond symptoms and looking for determinants of suffering, and support users to express their feelings. There are also recommendations regarding personalization options, feedback and continuous evaluation, ethics, accessibility, privacy and security.

Table 3 presents the categories and recommendation titles. Recommendations were numbered to facilitate their use and the numbers do not express priority. They are listed and explained following:

1. Allow the user to indicate persons of reference who should be contacted in the event of a crisis. People with depression, in a moment of a crisis, may attempt suicide and other harmful actions. At this point, it is dangerous to rely only in software solutions. The system may previously, preferably during first interaction, asks for contact information of persons of reference that had previously consented to be contacted in a crisis situation.

2. Provide shortcuts so that persons of reference are quickly reached out to in case of a crisis. Once the system has the contact of persons of reference, the user interface should offer quick access to them. For instance, offering a shortcut for calling the person with one click.

3. Offer information for human specialized care, preferably free or low-cost. Some users may rely on chatbots because they do not have easy access to human specialized care. If available, the system may support the users to find human specialized care that are accessible for them. This is possible in institutions, as universities or companies, but also pointing to low-cost healthcare or governmental services.

4. Go beyond the symptoms. Identify the determinants of suffering. While talking to human therapists, it is common to start the conversation mentioning symptoms, like difficulty sleeping or poor concentration. However, efficient treatment would go beyond the symptoms and look for the determinants of suffering as discrimination or financial problems. The chatbot should not stop conversation after finding first symptoms.

5. Check which protective factors have already been tried. As symptoms of depression are generally related to undergoing determinants of suffering, patients may have been ill for a period of time and already tried other solutions. The interaction with the chatbot may be more efficient if the user was asked to report protective factors that have already been tried.

6. Intervention proposals must cover self-oriented but also psychosocial solutions, at different levels of care. After identifying the determinants of suffering, the chatbot must be able to suggest interventions oriented to the individual and of localized actions, as sleep hygiene or anxiety reduction techniques, but also those that may require institutional and social support, such as looking for financial assistance or support groups.

7. Provide interaction options for uncertain subjective feeling. Users may face difficulties in expressing how they are feeling clearly. Therefore, create different strategies to infer the user's emotional state and design for moving further in the conversation. Data from sensors, behavior on social networks, medical history, and academic literature may be used to illustrate and support conversation.

8. Do not add pressure on the user. Although lethargy is one of the symptoms of depression, anxiety is also a recurring symptom. Interaction options that make pressure, demand quick answers, or keep asking the same question may worsen the user's emotional state. The chatbot should be designed to try different strategies to keep the conversation going, but also to wait if necessary, not rushing nor guilting the user.

9. Use emojis. Emojis are widely used in text conversations nowadays and add a more natural tone to chatbot communications. Choose emojis that are not ambiguous and that demonstrate empathy and care.

10. Use icons. Current chatbot technologies are able to chat about different topics. However, better performance is reached when the chatbot previously knows the conversation topic and/or was developed to that aim. In this sense, icons may be added to the user's interface to support quick identification of chat topics or intervention categories.

11. Inform the estimated time for the dialogue. Users may want to chat frequenty as part as their routine. In this case, to know in advance the estimated period of time for a conversation may support users to plan and adhre to frequent conversations. Although this information should be available, it should not sound as an obligation to finish the conversation in that specific period of time.

12. Adopt lightness in the dialogue. Chatbot should avoid technical terms, as "dysthymia" or "lethargy", or strong sentences as with "you should" or "never do that". On the other hand, the chatbot may ask clear and gentle questions and let the user talk. Examples can be used with the user to clarify possible expected answers or similar situations.

13. Adopt empathy strategies. Design the chatbot to show interest in what the user has to say. Avoid judging, "imposing the truth", and quickly delivering interventions about what has to be done. Ask questions and express respect to the user's values, opinions and feelings. Try not to finish the conversation abruptly.

14. Offer options to personalize the chatbot. Some people feel more comfortable if they talk about personal issues to other people with same gender or age. Although these characteristics do not truly apply to a chatbot, some users may feel more comfortable if they can personalize the chatbot avatar, gender and speech style (e.g. use of slang).

15. Offer options for user naming and reference. To turn the conversation more personal, the chatbot can directly refer to

Table 3: Design recommendations for chatbots to support people with depression.

Depression oriented design recommendations	Crisis management	1. Allow the user to indicate persons of reference who should be contacted in the event of a crisis.
		2. Provide shortcuts so that persons of reference are quickly reached out to in case of a crisis.
		3. Offer information for human specialized care, preferably free or low-cost.
	Determinants and interventions	4. Go beyond the symptoms. Identify the determinants of suffering.
		5. Check which protective factors have already been tried.
		6. Intervention proposals must cover self-oriented but also psychosocial solutions, at different levels of care.
	Feeling	7. Provide interaction options for uncertain subjective feeling.
		8. Do not add pressure on the user.
General mental health design recommendations	Interface elements	9. Use emojis.
		10. Use icons.
		11. Inform the estimated time for the dialogue.
	Speech pattern	12. Adopt lightness in the dialogue.
		13. Adopt empathy strategies.
	Personalization	14. Offer options to personalize the chatbot.
		15. Offer options for user naming and reference.
		16. Offer options to pause and restart the conversation.
	Feedback	17. Ask for explicit feedback regarding the user's mental health.
		18. Ask for explicit feedback regarding the interaction.
		19. Inform the user about any data collection, storage and use.
	Acessibility	20. Go beyond physical impairments. Consider economic, language and litteracy differences.
		21. Design for interaction without internet connection.
	Privacy	22. Design for non-embarrassed, non-identifiable alert messages.
		23. Offer login with password.
	Security	24. Adopt security technologies to avoid undesirable access to personal health data.

the user by name, nickname or other term of reference the user prefers. Asking for this term during first conversations, may allow the chatbot to personaly refer to the user time to time.

16. Offer options to pause and restart the conversation. Users may need to interrupt the conversation for several reasons, including be called to do other activities or even to hide the conversation from other people. It could be frustrating and disappointing to have to say everything again to reach the end of the conversation. Add options to pause and catch up later on.

17. Ask for explicit feedback regarding the user's mental health. There are several ways to infer if the user is feeling better, but asking him/her is one of the most effective. Ask how the user is feeling in the beginning of the conversation, but also ask about his/her mental health at the end of the conversation. Guide the user accordingly. Check if interventions offered were useful.

18. Ask for explicit feedback regarding the interaction. Chatbots for mental health should be frequently checking if the conversational flow is adequate and supporting the users to recover. Ask the user to judge the interaction strategies and resources. Improve your solution on the fly or in a future release.

19. Inform the user about any data collection, storage and use. This is an ethical requirement to any research project or technical solution. However, while using chatbots that allow the user to freely text or speak, unpredictable data can be collected. If these data are being storage, transmitted or processed someway, the user should be aware of that. The user can always be remembered about that time to time in a natural way.

20. Go beyond physical impairments. Consider economic, language and litteracy differences. Accessible interaction for physical disabilities should be offered as screen reader compliant. However, design the chatbot considering that users may have different economic incomes (e.g. private healthcare may not be an option) or geographic limitations (healthcare services may not be fully available in a particular geographic location). Also design for different language skills and literacy levels.

21. Design for interaction without internet connection. Internet connection may be lost for several different reasons, even in technological centers. The abrupt interruption in a conversation about sensible issues may serve as a mental health trigger. Design to keep the conversation, at least for a while, and direct it to a smooth ending in the case of internet lost.

22. Design for non-embarrassed, non-identifiable alert messages. Stigma about mental health treatments prevents some people to look for help and adhere to treatments. Moreover, chatting with a machine can also sound weird to some people. Therefore, users may want to keep the interaction with the chatbot discreet. Alert messages that can be seen on a user's screen in a non-interaction time should be carefully designed to avoid showing personal data and/or association with mental health issues.

23. Offer login with password. Health data is sensitive and should be accessed only with the user's consent. Offer the option to login with password may avoid undesirable access to personal information and supports privacy.

24. Adopt security technologies to avoid undesirable access to personal health data. Data informed to and/or collected by the chatbot may be stored, processed and transmitted. They could be a target for a cyberattack aiming at undesirable expose. Consider the use of security technology to mitigate the risk.

5 CONCLUSIONS

This paper presents a set of 24 recommendations for the design of chatbots that aim to support people with depression. These recommendations were formalized by learning from a Benchmark of six available systems and from discussions with mental healthcare and HCI professionals. This research results aim to contribute to the rationale of these systems, in a human-centered design approach, putting the focus on aspects such as the user's mental health safety, needs for personalization, privacy, and security concerns.

It is worth mentioning that one strong aspect of this work relies on the involvement of six highly qualified specialists with academic and practical experience in the mental healthcare field and two HCI specialists with experience in neurodiverse and well-being. This multidisciplinary approach has been highlighted in the literature as needed for this type of work. One limitation of this work is that the recommendations were not totally applied in the design of a chatbot for this aim, but this is a work in progress at this moment.

Mental health is a very broad field and its pathologies have characteristics that need to be treated individually. Therefore, we carried out this study focusing on a specific pathology, depression. The application of the recommendations, as well as an analysis of whether they may be suitable for other contexts and pathologies, can be considered for future research.

6 ACKNOWLEDGMENTS

We acknowledge all the researchers of the AMIVE project and all specialists who joined the discussions. We also thank FAPESP and CAPES for financial support.

FAPESP - Process nº 2020/05157-9, *Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)*.

This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES)* - Finance Code 001.

REFERENCES

[1] Marcelo José Siqueira Coutinho de Almeida. 2013. Desenvolvimento de Benchmarks para sistemas multiagente: o caso da patrulha orientada a eventos. Retrieved May 5, 2022 from https://repositorio.ufpe.br/handle/123456789/12276 Publisher: Universidade Federal de Pernambuco.

[2] American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed. ed.). Autor, Washington, DC.

[3] Timothy W. Bickmore, Laura M. Pfeifer, Donna Byron, Shaula Forsythe, Lori E. Henault, Brian W. Jack, Rebecca Silliman, and Michael K. Paasche-Orlow. 2010. Usability of Conversational Agents by Patients with Inadequate Health Literacy: Evidence from Two Clinical Trials. *Journal of Health Communication* (2010), 197–210. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10810730.2010.499991.

[4] Timothy W. Bickmore, Daniel Schulman, and Candace Sidner. 2013. Automated interventions for multiple health behaviors using conversational agents. *Patient Education and Counseling* 92, 2 (Aug. 2013), 142–148. https://doi.org/10.1016/j.pec.2013.05.011

[5] GBD 2019 Mental Disorders Collaborators et al. 2022. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Psychiatry* (2022). https://doi.org/10.1016/S2215-0366(21)00395-3

[6] Rik Crutzen, Gjalt-Jorn Y. Peters, Sarah Dias Portugal, Erwin M. Fisser, and Jorne J. Grolleman. 2011. An Artificially Intelligent Chat Agent That Answers Adolescents' Questions Related to Sex, Drugs, and Alcohol: An Exploratory Study. *Journal of Adolescent Health* 48, 5 (May 2011), 514–519. https://doi.org/10.1016/j.jadohealth.2010.09.002 Publisher: Elsevier.

[7] Venkata Duvvuri, Qihan Guan, Swetha Daddala, Mitch Harris, and Sudhakar Kaushik. 2022. Predicting Depression Symptoms from Discord Chat Messaging Using AI Medical Chatbots. In *6th International Conference on Machine Learning and Soft Computing*. ACM, Haikou China, 111–119. https://doi.org/10.1145/3523150.3523168

[8] Terry Ellis, Nancy K. Latham, Tamara R. DeAngelis, Cathi A. Thomas, Marie Saint-Hilaire, and Timothy W. Bickmore. 2013. Feasibility of a Virtual Exercise Coach to Promote Walking in Community-Dwelling Persons with Parkinson Disease. *American Journal of Physical Medicine & Rehabilitation* 92, 6 (June 2013), 472–485. https://doi.org/10.1097/PHM.0b013e31828cd466

[9] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. (2017). https://doi.org/10.2196/mental.7785 JMIR Mental Health, Toronto, Canada.

[10] Vinícius F Galvão, Cristiano Maciel, and Ana Cristina Bicharra Garcia. 2019. Creating chatbots to talk with humans: HCI evaluations and perspectives. In *Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems (IHC '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3357155.3358460

[11] Raman Goel, Sachin Vashisht, Armaan Dhanda, and Seba Susan. 2021. An Empathetic Conversational Agent with Attentional Mechanism. In *2021 International Conference on Computer Communication and Informatics (ICCCI)*. 1–4. https://doi.org/10.1109/ICCCI50826.2021.9402337 ISSN: 2329-7190.

[12] Eric P. Green, Yihuan Lai, Nicholas Pearson, Sathyanath Rajasekharan, Michiel Rauws, Angela Joerin, Edith Kwobah, Christine Musyimi, Rachel M. Jones, Chaya Bhat, Antonia Mulinge, and Eve S. Puffer. 2020. Expanding Access to Perinatal Depression Treatment in Kenya Through Automated Psychological Support: Development and Usability Study. (Oct. 2020), e17895. https://doi.org/10.2196/17895 JMIR Formative Research, Toronto, Canada.

[13] Gaurav Kumar Gupta and Dilip Kumar Sharma. 2021. Depression Detection on Social Media with the Aid of Machine Learning Platform: A Comprehensive Survey. In *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*. 658–662. https://doi.org/10.1109/INDIACom51348.2021.00116

[14] Becky Inkster, Shubhankar Sarda, and Vinod Subramanian. 2018. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR mHealth and uHealth* 6, 11 (Nov. 2018), e12106. https://doi.org/10.2196/12106

[15] Brian Jack, Timothy Bickmore, Megan Hempstead, Leanne Yinusa-Nyahkoon, Ekaterina Sadikova, Suzanne Mitchell, Paula Gardiner, Fatima Adigun, Brian Penti, Daniel Schulman, and Karla Damus. 2015. Reducing Preconception Risks Among African American Women with Conversational Agent Technology. *The Journal of the American Board of Family Medicine* 28, 4 (July 2015), 441–451. https://doi.org/10.3122/jabfm.2015.04.140327 Publisher: American Board of Family Medicine Section: Original Research.

[16] Lizy Kurian John and Lieven Eeckhout (Eds.). 2017. *Performance Evaluation and Benchmarking*. CRC Press, Boca Raton. https://doi.org/10.1201/9781315220505

[17] Alita Joyce. 2020. *7 Steps to Benchmark Your Product's UX*. https://www.nngroup.com/articles/product-ux-benchmarks/

[18] Robert Boxwell Jr. 1996. *Vantagem Competitiva Atraves Do Benchmarking*. Mcgraw Hill - Import, São Paulo.

[19] Ahmet Baki Kocaballi, Shlomo Berkovsky, Juan C. Quiroz, Liliana Laranjo, Huong Ly Tong, Dana Rezazadegan, Agustina Briatore, and Enrico Coiera. 2019. The Personalization of Conversational Agents in Health Care: Systematic Review. (Nov. 2019), e15360. https://doi.org/10.2196/15360 Journal of Medical Internet Research, Toronto, Canada.

[20] Rafal Kocielnik, Raina Langevin, James S. George, Shota Akenaga, Amelia Wang, Darwin P. Jones, Alexander Argyle, Callan Fockele, Layla Anderson, Dennis T. Hsieh, Kabir Yadav, Herbert Duber, Gary Hsieh, and Andrea L. Hartzler. 2021. Can I Talk to You about Your Social Needs? Understanding Preference for Conversational User Interface in Health. In *CUI 2021 - 3rd Conference on Conversational User Interfaces (CUI '21)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3469595.3469599

[21] Jeya Amantha Kumar. 2021. Educational chatbots for project-based learning: investigating learning outcomes for a team-based design course. *International Journal of Educational Technology in Higher Education* 18, 1 (Dec. 2021), 65. https://doi.org/10.1186/s41239-021-00302-w

[22] Liliana Laranjo, Adam G. Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie Y. S. Lau, and Enrico Coiera. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association: JAMIA* 25, 9 (Sept. 2018), 1248–1258. https://doi.org/10.1093/jamia/ocy072

[23] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376175

[24] Qiao Ying Leong, Shreya Sridhar, Agata Blasiak, Xavier Tadeo, GeckHong Yeo, Alexandria Remus, and Dean Ho. 2022. Characteristics of Mobile Health Platforms for Depression and Anxiety: Content Analysis Through a Systematic Review of the Literature and Systematic Search of Two App Stores. *Journal of medical Internet research* 24, 2 (2022), e27388.

[25] Yi Joy Li and Hao Irene Luo. 2021. Depression Prevention by Mutual Empathy Training: Using Virtual Reality as a Tool. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 60–63. https://doi.org/10.1109/VRW52623.2021.00017

[26] Gale M. Lucas, Albert Rizzo, Jonathan Gratch, Stefan Scherer, Giota Stratou, Jill Boberg, and Louis-Philippe Morency. 2017. Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers. *Frontiers in Robotics and AI* 4 (2017). https://www.frontiersin.org/article/10.3389/frobt.2017.00051

[27] Joyce H. L. Lui, David K Marcus, and Christopher T. Barry. 2017. Evidence-Based Apps? A Review of Mental Health Mobile Applications in a Psychotherapy Context. *Professional Psychology: Research and Practice* 48 (2017), 199–210.

[28] Eri Maeda, Akane Miyata, Jacky Boivin, Kyoko Nomura, Yukiyo Kumazawa, Hiromitsu Shirasawa, Hidekazu Saito, and Yukihiro Terada. 2020. Promoting

fertility awareness and preconception health using a chatbot: a randomized controlled trial. *Reproductive BioMedicine Online* 41, 6 (Dec. 2020), 1133–1143. https://doi.org/10.1016/j.rbmo.2020.09.006 Publisher: Elsevier.

[29] Stanley John Mierzwa, Samir Souidi, Terry Conroy, Mohammad Abusyed, Hiroki Watarai, and Tammy Allen. 2019. On the Potential, Feasibility, and Effectiveness of Chat Bots in Public Health Research Going Forward. *Online Journal of Public Health Informatics* 11, 2 (Sep. 2019). https://doi.org/10.5210/ojphi.v11i2.9998

[30] Madison Milne-Ives, Caroline de Cock, Ernest Lim, Melissa Harper Shehadeh, Nick de Pennington, Guy Mole, Eduardo Normando, and Edward Meinert. 2020. The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review. (2020). Journal of Medical Internet Research, Toronto, Canada. https://www.jmir.org/2020/10/e20346.

[31] Adam S. Miner, Arnold Milstein, Stephen Schueller, Roshini Hegde, Christina Mangurian, and Eleni Linos. 2016. Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health. *JAMA Internal Medicine* 176, 5 (May 2016), 619–625. https://doi.org/10.1001/jamainternmed.2016.0400

[32] Pritika Parmar, Jina Ryu, Shivani Pandya, João Sedoc, and Smisha Agarwal. 2022. Health-focused conversational agents in person-centered care: a review of apps. *NPJ digital medicine* 5, 1 (Feb. 2022), 21. https://doi.org/10.1038/s41746-022-00560-6

[33] Pierre Philip, Jean-Arthur Micoulaud-Franchi, Patricia Sagaspe, Etienne De Sevin, Jérôme Olive, Stéphanie Bioulac, and Alain Sauteraud. 2017. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. *Scientific Reports* 7, 1 (Feb. 2017), 42656. https://doi.org/10.1038/srep42656 Number: 1 Publisher: Nature Publishing Group.

[34] Raquel Planas and Oriol Yuguero. 2021. Technological prescription: evaluation of the effectiveness of mobile applications to improve depression and anxiety. Systematic review. *Informatics for Health and Social Care* 46, 3 (2021), 273–290.

[35] Simon Provoost, Ho Ming Lau, Jeroen Ruwaard, and Heleen Riper. 2017. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. *Journal of Medical Internet Research* 19, 5 (May 2017), e151. https://doi.org/10.2196/jmir.6553

[36] Chengcheng Qu, Corina Sas, Claudia Daudén Roquet, and Gavin Doherty. 2020. Functionality of Top-Rated Mobile Apps for Depression: Systematic Search and Evaluation. *JMIR Ment Health* 7, 1 (24 Jan 2020), e15321. https://doi.org/10.2196/15321

[37] Anand Singh Rajawat, Romil Rawat, Kanishk Barhanpurkar, Rabindra Nath Shaw, and Ankush Ghosh. 2021. Chapter Five - Depression detection for elderly people using AI robotic systems leveraging the Nelder–Mead Method. In *Artificial Intelligence for Future Generation Robotics*, Rabindra Nath Shaw, Ankush Ghosh, Valentina E. Balas, and Monica Bianchini (Eds.). Elsevier, 55–70. https://doi.org/10.1016/B978-0-323-85498-6.00006-X

[38] Margherita Rampioni, Vera Stara, Elisa Felici, Lorena Rossi, and Susy Paolini. 2021. Embodied Conversational Agents for Patients With Dementia: Thematic Literature Analysis. *JMIR mHealth and uHealth* 9, 7 (July 2021), e25381. https://doi.org/10.2196/25381

[39] Natasha Randall, Selma Šabanović, and Wynnie Chang. 2018. Engaging Older Adults with Depression as Co-Designers of Assistive In-Home Robots. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '18)*. Association for Computing Machinery, New York, NY, USA, 304–309. https://doi.org/10.1145/3240925.3240946

[40] Damian F Santomauro, Ana M Mantilla Herrera, Jamileh Shadid, Peng Zheng, Charlie Ashbaugh, David M Pigott, Cristiana Abbafati, Christopher Adolph, Joanne O Amlag, Aleksandr Y Aravkin, et al. 2021. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet* 398, 10312 (2021), 1700–1712.

[41] Victon Malcolm R. Santos, Francisco Petrônio A. Medeiros, Heremita B. Lira, and Nadja Nóbrega Rodrigues. 2021. Benchmark Application for Scenario Analysis in the Educational Chatbots Development. In *2021 XVI Latin American Conference on Learning Technologies (LACLO)*. 302–309. https://doi.org/10.1109/LACLO54177.2021.00038

[42] Maria J Serrano-Ripoll, Rocío Zamanillo-Campos, Maria A Fiol-DeRoque, Adoración Castro, and Ignacio Ricci-Cabello. 2022. Impact of Smartphone App–Based Psychological Interventions for Reducing Depressive Symptoms in People With Depression: Systematic Literature Review and Meta-analysis of Randomized Controlled Trials. *JMIR mHealth and uHealth* 10, 1 (2022), e29621.

[43] Michael J. Spendolini. 1992. *The Benchmarking Book*. Amacom Books, New York.

[44] Katarzyna Stawarz, Chris Preist, Debbie Tallon, Nicola Wiles, and David Coyle. 2018. User Experience of Cognitive Behavioral Therapy Apps for Depression: An Analysis of App Functionality and User Reviews. *J Med Internet Res* 20, 6 (06 Jun 2018), e10120. https://doi.org/10.2196/10120

[45] Colleen Stiles-Shields, Enid Montague, Mary J. Kwasny, and David C. Mohr. 2019. Behavioral and cognitive intervention strategies delivered via coached apps for depression: Pilot trial. *Psychological Services* 16, 2 (May 2019), 233–238. https://doi.org/10.1037/ser0000261

[46] Colleen Stiles-Shields, Enid Montague, Emily G. Lattie, Stephen M. Schueller, Mary J. Kwasny, and David C. Mohr. 2017. Exploring User Learnability and Learning Performance in an App for Depression: Usability Study. *JMIR human factors* 4, 3 (Aug. 2017), e18. https://doi.org/10.2196/humanfactors.7951

[47] John Torous, Sandra Bucci, Imogen H Bell, Lars V Kessing, Maria Faurholt-Jepsen, Pauline Whelan, Andre F Carvalho, Matcheri Keshavan, Jake Linardon, and Joseph Firth. 2021. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry* 20, 3 (2021), 318–335.

[48] Md Zia Uddin, Kim Kristoffer Dysthe, Asbjørn Følstad, and Petter Bae Brandtzaeg. 2022. Deep learning for prediction of depressive symptoms in a large textual dataset. *Neural Computing and Applications* 34, 1 (Jan. 2022), 721–744. https://doi.org/10.1007/s00521-021-06426-4

[49] Francisco Albernaz Machado Valério, Tatiane Gomes Guimarães, Raquel Oliveira Prates, and Heloisa Candello. 2018. Chatbots Explain Themselves: Designers' Strategies for Conveying Chatbot Features to Users. *Journal on Interactive Systems* 9, 3 (Dec. 2018). https://doi.org/10.5753/jis.2018.710 Number: 3.

[50] Caroline Wachtler, Amy Coe, Sandra Davidson, Susan Fletcher, Antonette Mendoza, Leon Sterling, and Jane Gunn. 2018. Development of a Mobile Clinical Prediction Tool to Estimate Future Depression Severity and Guide Treatment in Primary Care: User-Centered Design. *JMIR mHealth and uHealth* 6, 4 (April 2018), e95. https://doi.org/10.2196/mhealth.9502

[51] Robert Whitaker. 2017. *Anatomia de uma Epidemia: pílulas mágicas, drogas psiquiátricas e o aumento assombroso da doença mental*. Fiocruz, Rio de Janeiro.

[52] WHO. 2021. *World Health Organization - Depression*. Retrieved May 3, 2022 from https://www.who.int/news-room/fact-sheets/detail/depression