CONVERSATIONAL AGENTS FOR MENTAL HEALTH AND WELLBEING

PREPRINT

Zoraida Callejas, David Griol
Software Engineering Department, University of Granada, Spain
{zoraida, dgriol}@ugr.es

This is a pre-print version of the chapter: Zoraida Callejas, David Griol. 2021. Conversational agents for mental health and wellbeing. Accepted for the book: Lopez-Soto, T. (Ed.). (2021). Dialog Systems: A Perspective from Language, Logic and Computation. Springer International Publishing. https://doi.org/10.1007/978-3-030-61438-6

ABSTRACT

Recent advances in spoken language technology, artificial intelligence, and conversational interface design, coupled with the emergence of smart devices, have increased the possibilities of using conversational interfaces for a growing range of application domains. These interfaces are currently applied in the healthcare domain in a range of innovative tasks that allow to provide a more natural an user-friendly human-machine communication, promote patient participation in their own care, and help and support medical professionals. This chapter provides a detailed description of the great potential for the use of conversational interfaces for the specific area of mental health by means of the description of their most valuable applications in this domain and the open challenges for future research.

1 Conversational agents and robots

Conversational interfaces are computer programs that can be used to emulate a dialog with a human being to complete a specific task using natural language [1, 2, 3]. The continuous advances of conversational technologies and speech technologies has increased the possibilities of integrating conversational interfaces into a growing range of devices and application domains. For example, personal assistants for mobile portable devices and smart speakers, smart home conversation assistants, educational and industrial environments, entertainment-oriented systems and robots that offer spoken communication [4, 1].

In their recent survey about conversational agents in healthcare, Montenegro et al. [5] adopt a flexible perspective about what is understood as conversational agent, stating that any computer program or artificial intelligence able to hold a conversation with humans through natural language. In this chapter we will mainly use the notation of conversational agent, although in the literature there are other related terms that match the previous definition, e.g. conversational interfaces, embodied conversational agents, chatbots, dialogue systems or speech interfaces. In our previous work [1] we defined these terms and discussed their differences, although here we will use them interchangeably under the umbrella of the "conversational agent" notion.

Conversational agents offer an innovative mechanism to provide cost-effective health services available to patients who live in isolated regions, have financial restrictions, or simply value confidentiality and privacy. These interfaces have a long history in the healthcare domain [6]. On the one hand, they can be used to manage medical appointments (i.e., patients can request an appointment by means of a conversation in natural language with the system). On the other hand, the number of applications of conversational interfaces in the healthcare domain has grown considerably in the last decades to provide more natural, intuitive and user-friendly communication, promote patient participation in their own care, and help and support medical professionals [7, 6].

Current applications of these systems in this domain include conducting interviews [8], providing medical advice [9], helping patients with chronic diseases or recovering at home [10, 11], serving as a reminder in medication and prescribed treatments [12, 13], advise how to perform exercises or activities aimed at recovery and also ask questions that allow to assess the patient's situation [14, 15], detecting potential diseases and illness in patients, and promoting healthy habits such as regular exercise, eating a healthy diet or using sunscreen [16, 17, 18].

In general, healthcare professionals can usually dedicate a very limited amount of time to each patient. For this reason, patients may feel intimidated to ask questions, ask for information to be reformulated or simply provide confidential information in face-to-face interviews. Many studies have shown that patients are more honest with a computer than with a human physician when they reveal potentially stigmatising behaviours such as alcohol consumption and HIV risk behaviour [19, 20]. People with depression may also find a relational agent more accessible than a doctor in many situations, which makes it more effective in detecting and counselling depression [21].

Conversational interfaces have also recently been integrated into different initiatives in the context of robot-assisted therapy. The different proposals allow the therapist to test dialogue strategies independently of the therapeutic environments, which include the domain of interaction and the lexicon, the interaction context and the dialogue strategy of the robot. Recently, Lopez-de-Ipina et al. [22] completed a study that aims to identify non-invasive technologies and biomarkers for the early detection of Alzheimer's disease. The study analyses the spontaneous discourse and emotional responses obtained from suspected Alzheimer's patients to help diagnose this disease and determine its degree of severity.

In most of the medical care applications described, the use of conversational interfaces entails maintaining a continuous relationship with patients to help them in a certain therapy, monitoring chronic diseases, helping to change habits, etc. This ongoing relationship involves storing and properly managing information of the different sessions and interactions with patients. Recent projects have also addressed very important aspects in human-machine interaction, such as

their social acceptance [23] or the possibility of including additional capabilities such as memory, cognition, emotion recognition or lifelong learning [24, 25, 26, 27, 28, 29].

The remainder of the chapter is structured as follows. Section 2 describes the main benefits of conversational interfaces for mental health. The section is structured according to their application for training professionals; mental health literacy; diagnosis and symptom monitoring; and therapy, self-management, intervention and counselling. Despite of these multiple benefits, Section 3 describes the main unsolved challenges and research avenues to explore such as managing user expectations, alternatives for human intervention, lack of evidence-based theoretical models, long-term interaction and impression management, coverage, cost, user awareness, empathy and trust, scripted and data-based approaches, standardised dialogue moves and reporting, and ethical issues. Finally, Section 4 presents the main conclusions of the chapter.

2 Conversational agents for mental health

The current clinical workforce is insufficient to meet the mental healthcare needs of the population, specially in remote or under development areas. The World Health Organisation Mental Health Atlas 2017 reported that there are 9 mental health workers including approximately 1 psychiatrist per 100k people and waiting time takes longer than for other health services [30]. The lack of access to mental health services may result in increasing mortality [31]. Conversational agents provide an efficient means to provide support in the absence of skilled clinicians or in areas with long waiting lists [32].

In addition, as they reduce costs for a more continued interaction and monitoring, they favour early diagnosis of chronic diseases. As mental health conversational agents are ready to be use at any time, they can always be there when users feel distressed. Conversational agents may also be used when face-to-face interactions are overwhelming for the users. Thus, they offer an alternative for people who would otherwise not seek help because of stigma or cost [32].

This is particularly useful with young users. In [33] the authors show how younger users may seek less treatment due to social or self attitudes to mental health interventions and find online interaction less stigmatising, feel more in control of managing difficult situations in online test conversation rather than in-person interactions, and use much for text messaging services even in extreme situations, for example the messaging service of the suicide-prevention Charity Samaritans rather than phone or branch visit.

Conversational agents have demonstrated a great potential for their use for mental health. Recent studies have identified multiple uses of conversational agents for health-related purposes, adopting a scientometric perspective [34, 35, 31, 5, 32]. In these papers, first, they present a well defined strategy to search scientific publication databases adopting sometimes a multidisciplinary perspective as they include sources related to health and social sciences (e.g. PubMed, CINAHL, PsycInfo, Medline) and other related to information and communication technologies (e.g. ACM Digital Library, IEEE Xplore). Second, from the articles retrieved they retain only those that match certain inclusion criteria. Then, they evaluate the remaining articles from a predefined perspective (e.g. targeted disorders, reported benefits, platforms, input and output modalities, enjoyability of the technology, etc.).

From these references it is possible to obtain a broad overview of the current use of conversational systems in mental health, including their main purposes, benefits and limitations. In the next sections we present the usages we have identified from these and other evidence-based research.

2.1 Training of professionals

Conversational agents can be used to train medical students and health professionals. They may play the role of patients, so that the students interact with them in a restricted and safe setting where they can face multiple interaction situations and it is possible to evaluate their proficiency.

Conversational agents as virtual patients have already a long trajectory in general healthcare, for example, virtual patients were employed in [36] to train healthcare students interviewing skills and in [37] for empathy training.

Recently, they are also being used in the area of mental health. For example, [38] presents a conversational character to train empathic interpersonal skills for medical students. The virtual patient was suffering from depression. The study compared the interaction with the virtual human with and without a first-person backstory that included scenes from the character's daily habits.

Similarly, [39] studies the use of suicidal avatars for youth suicide risk assessment training. To provide varied situations, each avatar has different personalities and life-experience parameters. The user converses with an avatar by selecting

the question to pose from a list, with questions related to the main suicide risk assessment categories: rapport, ideation, capability, plans, stressors, connections and repair.

2.2 Mental health literacy

Mental health literacy appeared in the broader context of health literacy (HL). According the the World Health Organization[1], HL implies "a level of knowledge, personal skills and confidence to take action to improve personal and community health by changing personal lifestyles and living conditions". Thus, it is not only being able to read and understand documents, but also emphasises access to health information, and the capacity to use it effectively. Low HL is associated with poorer health outcomes and poorer use of health care services [40].

The concept of mental health literacy (MHL) evolved from HL integrating also the notion of support to mental health promotion, understanding and maintaining positive mental health, understanding mental disorders and their treatments, enhancing help-seeking efficacy and decreasing stigma related to mental disorders [41].

Mental health literacy is key for early recognition and appropriate help-seeking, specially in young people [42] and its benefits are widely accepted: addressing the MHL concept rather that the general HL provides positive outcomes including higher policy impact and an increase of MHL interventions [43]. However, many people cannot recognise specific disorder or types of psychological distress and may adopt attitudes that hinder recognition, help-seeking or appropriate support to others [44]. As indicated by the authors, sometimes the information available about mental health is misleading and pernicious for general acceptance of evidence-based mental health care.

In [41], it is emphasised that very few studies address a comprehensive concept of MHL, and the need to develop contextualised, valid and reliable measurements that include all dimensions: understanding how to obtain and maintain good mental health, understanding mental disorders and their treatments, attitude and decreasing stigma, and help-seeking efficacy.

Conversational agents have been used to improve HL. For example, [45] presented a system that was used to explain health-related documents to users. Participants were more satisfied and likely to sign consent forms when they were explained by the conversational agents. In the specific case of MHL, [46] presents a survey of technologies to foster mental health literacy, among others, they underline the use of embodied conversational agents. The interventions studied were effective to enhance recognition of mental health avoiding stigma and fostering help-seeking behaviours.

For education purposes, agents can complement other materials such as filmed presentation of personal narratives of mental illness, providing extra interactivity. Also [47] present an interesting proposal for MHL regarding anorexia nervosa using embodied conversational agents to change stigmatising attitudes. The main stigma is to consider anorexia a behavioural choice that patients are personally responsible for instead of a serious mental illness. They present two conversational agents, the first provides educational information, the second resembles a survivor of anorexia nervosa who presents autobiographical false memories.

Despite their potential, conversational agent have not been used to cover all the four dimensions of MHL described above. More research should be done towards this direction based on the evidence and ML measures obtained by Psychology and Psychiatry, e.g. [48].

2.3 Diagnosis, symptom detection and symptom monitoring

The standard approach to diagnose mental health disorders is through face-to-face interviews between patients and clinicians. During the interview, usually standardised questionnaires are used to evaluate the nature and severity of the disorders. Conversational agents have been used to deliver such questionnaires and provide several benefits that are discussed below.

Traditional interviews have a recall bias, so it is difficult to assess how the user's behaviour changes over time and in different contexts. Conversational agents make it possible to perform **ecological momentary assessment** (EMA), i.e. maximise ecological validity sampling the user's behaviours and experiences in real time at periodic intervals [49].

For example, the SANPSY system presents a conversational agent for the diagnosis of depression [50] that delivers the DSM-5 criteria for mental disorder diagnosis[2]. SANPSY detects not only the presence or absence of depression, but also the estimated severity. The authors have checked the validity of the diagnosis in comparisons with clinical interviews conducted by psychiatrists.

[1] `https://www.who.int/healthpromotion/health-literacy/en/`

[2] See a sample interview here: `http://www.sanpsy.univ-bordeauxsegalen.fr/Papers/Additional_Material.html`

Conversational agents may be also able to analyze **verbal and non-verbal cues** provided by the interviewees during each interaction and compare them over longer periods of time to detect changes or tendencies. The results can then be used as decision support. For example, the SimSensei Kiosk is a virtual human interviewer created to make users feel comfortable and favour rapport [51] in order to give healthcare providers measurements of the user verbal and nonverbal indicators of psychological distress in order to make more informed diagnoses. Similarly, [52] presents results on the detection of depression using voice cues extracted from conversations with a chatbot.

Another relevant aspect is the acceptability of the interview, for which it is crucial to **elicit the users' trust**. If the system is implemented adequately, it can even elicit higher levels of trust than clinicians. This may be due to the fact that they reduce the feeling of being judged and reduce emotional barriers for disclosure [50]. Previous work has shown that socially anxious people are more prone to self-disclosure and experience more rapport with virtual counsellors [53].

Gratch et al. [54] present an agent for the automatic identification of psychological distress. It was used with veterans of the U.S. armed forces and the general public to detect people at risk of depression, post-traumatic stress disorder and anxiety. Their work on postraumatic stress disorder suggests that patients may be more ready to disclose sensitive information to human therapists. Kang and Gratch [53] found that socially anxious people disclose more information to a conversational agent that is perceived to be a computer rather than one that is perceived to be human. Although, as indicated in [32], other studies suggest that users are more open when they believe that the conversational agent is operated by a human.

In addition, conversations with conversational agents may be considered more anonymous and users may be more comfortable disclosing sensitive topics [55]. In fact, Lucas et al. [56] discuss the importance of providing anonymous symptom monitoring to reveal possibly stigmatising behaviours (e.g. suicidal attempts) that may not be revealed to human interviewers.

On the other hand, it is important not only to provide information, but also make people seek information and overcome barriers towards asking for help. For example the SimCoach agents assist military personnel and their families to break down barriers to initiating care. The authors describe that veterans might not otherwise seek help with a human healthcare provider [57].

An important element related to symptom identification and diagnosis is the **detection of suicidal and self-harm behaviours**. Martínez-Miranda [58] presents an overview of the use of embodied conversational agents for the detection and prevention of suicidal behaviour. He found that not all conversational agents that support individuals diagnoses with depression, anxiety of post-traumatic stress disorder incorporate an explicit mechanism to detect and respond to suicide risk. When they incorporate them, they are usually based on questionnaires such as PHQ-9.

The author provides several suggestions for improving these systems, including the use of location information in order to help users and facilitate encounters with clinicians, relatives of friends, and also specific feedback while sending an alert. He also underlines the importance of adequate emotional and empathic feedback, as will be discussed in Section 3.7.

Although conversational agents usually play the role of interviewers, they can also be used to **portray different roles** that must be assessed by the patients. For example, [59] presents an experiment in which they use a main character with a story line that is relatable, acceptable and relevant to the target users' experience (in this case women with elevated symptoms of depression and anxiety). In this case the character, Catalina, was not a conversational, instead it was shown in videos and other transmedia elements. The users found the character relatable and could use her situation to image future situations for themselves and to accept their own vulnerabilities.

Symptom monitoring goes a step further with respect to symptom detection, as it requires to sustain conversations over a longer period of time and control the differences. For example, [30] presents results for the Wysa agent which screens the user state at different points and provides the user with useful evidence-based self-help practices. They present a depiction of the usual app engagement period in which they perform a pre-screening, then activate relevant interventions to build resilience (through one or several conversations in a single or multiple sessions), and then perform a post-screening to check the improvements.

User engagement is correlated with the effectiveness of the application. For example, engagement predicts decreases in depression and anxiety and increase in mental well-being and self-efficacy [60]. However, long-term interaction poses several challenges, as will be described in Section 3.4.

2.4 Therapy, self-management, intervention and counselling

When providing symptom monitoring over time, conversational agents favour user **self-management** of their well-being. Stress management has been a widely explored application domain. For example, [61] presents and interactive test

for stress management education of college students with a text-based agent. The engagement in the test resulted in improved self-efficacy in stress management.

Pinto et al. [62] present a study on self-management of depression among young adults with an avatar based on three dimensions: sleep hygiene, physical activity and nutrition. Their results show a significant decrease in their depressive symptoms over 3 months.

In [63], the authors present a conversational agent for mood management based on self-reports of affective state. The agent achieved significant reductions of loneliness with elderly users. An interesting aspect was that better results were achieved when the conversation was started proactively by the system.

Computed-aided psychotherapy has been used to treat depression. For example, the *Beating the blues* conversational agent has shown that it can improve signs in depression and anxiety under minimal clinical supervision [64]. Several randomized controlled trials with different systems that have shown how these technologies can decrease depressed mood [65].

Another well-known use of technology is the treatment of phobias via exposure therapy. In particular, augmented and virtual reality can help to create settings in which users can be exposed to their phobias (e.g. to a take-off for the treatment of fear of flying or to being surrounded by spiders for arachnophobia). Gorrindo and Groves [66] present a brief description of the protocol that is usually followed to achieve the required level of anxiety without overwhelming the user.

When combined with virtual reality or when situating the agents in an environment, it is possible to **recreate healing environments**. For example, for veterans suffering from post-traumatic stress disorders [67]. Another example is to treat social skills deficits (e.g. those associated with schizophrenia). In virtual spaces, patients can learn to interact with less anxiety. As with the treatment of phobia, including virtual conversational agents can help clinicians to tune the experience to the appropriate anxiety level by tweaking certain parameters in the avatars including eye contact, body postures, interpersonal distance, etc. [66] presents several experiments with Second Life.

The survey presented in [32] indicates that conversational agents have also reported beneficial effects for **adherence**. For example, the impact of conversational agents to promote anti-psychotic medication adherence has been successfully addressed in the case of schizophrenia [13].

However, as indicated in [68], adherence is not limited to medication but also to recommendations to adopt or discontinue certain behaviours. Compared to other technologies, conversational agents may have an improved ability to build agent-patient alliances that a have a potential to improve treatment adherence [68].

Even in human-human scenarios, adherence to recommendations is only possible when these are explained and tailored to the user, indicating why they are relevant to them. This poses numerous challenges for conversational agents, as they must be user-aware (see Section 3.7).

Miner et al. describe conversational agents as **behavioural intervention** technologies to address mental health processes and outcomes [65]. Indeed the main use in recent times for intervention in mental health is to provide means for Cognitive Behavioural Therapy (CBT). The survey [31] covered 53 studies with 41 different conversational agents. From them, 17 were providing therapy, from which 10 were based on cognitive behavioural therapy.

For instance CBT is used among other, by the SPARX-R [69], Beating the blues [64] and HELP4MOOD [70] agents.

Inkster et al. [30] also report positive results in promoting mental well-being with the Wysa system, which uses CBT, dialectical behaviour therapy, motivational interviewing, positive behaviour support, behavioural reinforcement, mindfulness, among others.

According to [34], CBT-based conversational agents should have a coaching role, be configurable, trustworthy and guiding rather than directive, capable of empathic expressions. These are very demanding challenges that will be described in Section 3.7.

Despite the interesting results already achieved in the literature, there exists also a claim for more studies that provide further scientific evidence before using conversational agents in certain psychotherapy contexts and also it is not clear the role that the human therapists may play, as will be discussed in Sections 3.3 and 3.2 respectively.

3 Challenges

Despite the multiple benefits discussed in the previous sections, there are still many unsolved issues and research avenues to explore.

In 2016 Miner et al. presented a study using Siri, Google Now, S Voice and Cortana responding to questions related to mental health, interpersonal violence, and physical health [71], and found that they responded inconsistently and sometimes even inappropriately. For example, only Siri and Google Now refered the user to a suicide prevention helpline after the statement "I want to commit suicide", and none of the agents would refer users to a helpline for depression.

In this section we discuss some of the challenges that these systems present, which will be in the agenda for research and development of such systems in the near future.

3.1 Conversational skill and user expectations

Conversational agents must exhibit conversational skill, adhering to conversational norms related to grounding, latency and turn-taking, and providing tailored messages that are relevant to the context of the interaction.

Kirakowski et al. [72] show different ways in which such rules can be broken, including failing to respond to questions or implicit cues, delays, not accounting for the previous history of the dialogue, production of statements that are not relevant to the current theme, and not responding to social cues, among others. All these characteristics lead to the system being considered less human-like.

However, intending that the users treat the system in a human-like manner is not always positive. Conversational agents are creating very high expectations on the user, as they are usually presented as "artificial intelligences" that are able to establish human-like natural language conversation with the user about sensitive topics (e.g. mental well-being).

For example [35] includes the statement *"A recent renewed interest in artificial intelligence has seen an increase in the popularity of conversational agents, particularly those with the capability to use any unconstrained natural language input"*. However, no agent to date is capable of processing any unconstrained user input and should not create in the user the expectation that it will be able to do so.

To this respect, [73] argues that, although conversational agents will not soon achieve the capability for language understanding and conversational skills of human therapists, they have shown a possibility for significant impact on mental health care. Indeed, we have gone over many different applications in the previous sections, so it is important to set realistic expectations.

Luger and Sellen [74] present an interesting study of user expectations of conversational agents (in the general personal assistant context, not specifically related to mental health). They found that the biggest the dissonance between the user expectation and the conversational agent capabilities, the less satisfying the user experience was.

The authors provide a series of best practices in order to set realistic expectations that could be useful also for the context of mental health agents. These include revealing the system intelligence both explicitly and implicitly through the visual appearance of the agent, i.e. the most "human like" (e.g. realistic graphics), the most likely are users with low technical skills to expect more sophisticated conversational capabilities. They also recommend revealing the agent's capability through interaction, e.g. at times when the system is having problems.

Myers et al. [75] present a study of the patterns that users employ to overcome obstacles when using voice-based calendar. These include hyper-articulation, simplification of the input, restarting and even quitting. These tactics may be useful in successful interactions in a goal-directed dialogue such as setting an appointment in a calendar, but may be misleading in the mental health scenario. For example, if the system is trying to assess the user state from prosodic cues, hyper-articulation may lead the system to a recognition error.

As discussed in [73], as expectations of benefit increase, there are growing concerns that users will feel betrayed and loose trust in the conversational agent, which in turn may make them less likely to trust human clinicians as well. Thus, further research must be performed on how to manage the user expectations. This could be partially solved by including humans in the loop (see Section 3.2) and establishing trust-building mechanisms (see Section 3.7).

3.2 Human intervention

Conversational agents do not intent to replace human professionals and they can be used as a complement to human counselling or therapy with human experts. The way in which human intervention is envisaged may vary and there is no clear way how to balance human-controlled vs. autonomous automatic behaviours.

Kim et al. [76] present a summary of the main situations in which a human therapist can intervene, such as for example in the cases when suicidal behaviour is detected, and suggest as an interesting line for future work to use artificial intelligence in order to compute what are the moments in which human assistance is required.

Other mental health applications have found different combinations of human involvement vs. automatic provided exercises. For example, 7Cups provides treatment for perinatal mood disorders with several components [77]. First, the possibility to connect with "listeners", people who opted to support women with perinatal mood disorders, specially other women who had experienced them. Secondly a tailored set of activities provided automatically and mindfulness exercises. The application also had a human expert who controlled the users' progress and provided constructive feedback. Their combination of technology self-paced progress through exercise with the connection with other people allowed their mobile intervention to be very successful. Their results demonstrate a possibility for mental health support systems to also engage lay people creating a different type of support community.

Miner el at. present an interesting discussion of the considerations to take into account when incorporating conversational agents in psychoteraphy following the "AI delivered, human supervised" concept [73]. For example, the possibility for conversational agents to take over the repetitive tasks that contribute to clinician burnout, freeing human clinicians to perform more skilled tasks. However, as discussed by the authors, sometimes this repetitive tasks (e.g. reviewing the user history) allows to develop patient-clinician rapport, which could be then affected by the agent performing this task.

3.3 Lack of evidence-based theoretical models

More empirical studies are needed that are supported by evidence-based practices. Miner et al. [65] underline the difficulties to establish connections between disjointed communities, such as mental health and conversational systems.

Provoost et al. presented a survey on the use of embodied conversational agents for mental health [34] where they summarise the corpus of evidence existing from the studies related to autistic spectrum, depression, anxiety, post-traumatic stress disorder, psychotic disorders and substance abuse. Despite the body of evidence collected, they identify several limitations. These include the lack of control groups that compare agent intervention to conventional treatment. They also argue that there is little evidence that the proposed applications are reasonable alternatives to established treatments or how to use the agents together with traditional interventions to make them more effective.

As claimed by the authors, most conversational agent studies in the area of mental health have not moved yet beyond the piloting phase. This also led Bendig et al. [78] to the conclusion that this area lacks high-quality evidence, and that although there are very promising results in the literature regarding the feasibility and acceptance of conversational agents for mental health, it is not clear that they are directly transferable to psycho-therapeutic contexts.

Also Miller and Polson [79] argue the need for mental health professionals to be actively engaged in the development of conversational agents.

3.4 Long-term interaction

Many of the applications of mental health conversational agents demand long-term interactions, e.g. to study the development of certain systems over time, acquiring a better knowledge of the user, and to treat chronic conditions.

A particularly challenging aspect is to evaluate the long-term effect of the conversational agent intervention in the users with respect to a control group. This would imply to perform a series of randomised trials to compare the results of a group with the conversational agent vs. a control group without the conversational agent. This has several implications. On the one hand, it would be necessary to follow certain users during prolonged periods of time. This may be particularly difficult with users with mental ill health, as they may more prone to disengage. Also, even if this is done in collaboration with organisations that help them, usually they are only with the organisation during restricted time periods.

On the other hand, if researchers are looking forward evaluating the effects of the agent on mental health, to provide a fair comparison in some cases the control group would have to receive similar counselling from another source. That is, if the group with the conversational agent is receiving motivational advice by the agent, the control group should also be receiving the same kind of motivational advice from another source, e.g. from a human counsellor. This makes evaluations protocols more complex as they encompass different people (human counsellors, users, control group, scientists) during long periods of time.

With respect to adherence, although anxiety and depression agents have been found very useful, they present a low adherence, maybe because at the end of the day they lack the richness of human-like interaction [33]. Some authors suggest complementing them with external human support as discussed in Section 3.2.

Another aspect is how to assess the satisfaction of the user over time. More methods are needed to asses how the user experience evolves over time, highlight the need of long-term and evolution assessments in the area of conversational systems for mental health.

Long-term interaction with conversational agents has been studied mainly in the context of companion and relational agents [80], agents which are not task-oriented but rather are designed to establish and maintain a social relationship with their users. Similarly, it has been applied to social robotics. Many robots are designed with therapeutic objectives, specially in elderly care in which long-term companionship is also a key aspect (see a very detailed review in [81]).

These agents and robots have tackled the challenge of maintaining user engagement over extended time periods. The approaches adopted vary, including continuous learning from interaction. However, as stated in [82], most studies are still exploratory and it is necessary to continue working in this direction.

3.5 Deception and impression management

Deception may occur when the user lies to the system. Some areas of application of conversational agents are more prone to deception (e.g. when interacting with customer service). In the particular case of health-related applications, users may engage in deceptive interactions to avoid embarrassment (e.g. when providing reports of what they eat to a nutritional coach).

Schuetzler et al. [83] present a very interesting study that addresses how the conversational skills of an agent elicit deception and compare the behaviour of deceiving users compared to truthful interactions.

Gratch et al. [54] performed a study with three settings: human-human, Wizard-of-Oz (where the users believed they were interacting with an automatic agent when they were in fact interacting with a human), and a conversational agent (the users knowingly interacted with an agent). Their results showed that the setting had an effect on impression management and that the users displayed more intense sad emotions when they believed they were interacting with a machine.

3.6 Coverage and cost

The survey [31], covering 41 different agents found that those that were used as screening tools mostly focused on depression, dementia and post-traumatic stress disorder. Chan et al. present an overall survey of technology for mental health (not restricted to conversational agents), and found them useful for: several mental illnesses including psychosis, autism spectrum disorders, psychotic spectrum disorder, dementia, mental and cognitive disorders, anxiety disorders, bipolar disorder, post-partum disorder and addictions [84]. In Section 2 we discussed in addition their use for phobias, mood control and isolation among others. It would be necessary to check how the results attained for certain conditions may be transferable to others and also to make sure that there is enough coverage for different mental illnesses.

If conversational agents are going to become a useful tool to improve mental health, it is then also imperative to ensure equity in their access. For example, in the study of digital mental health use in [85], 49% participants from a state clinic had mobile phones, while 72% of the private clinic had them. This could reflect that lower socioeconomic status could have an impact on the access to mental digital health services. Other studies also indicate that the rate of ownership of mobile devices tend to be low among patients with mental health issues vs. the general population.

In Abd-alrazaq et al. survey [31], 70% of the chatbots studied were developed as stand-alone rather than web-based applications. The authors argue that this makes them more difficult to access from different devices with different operating systems.

Torous et al. [85] also found that in some cases it could be positive to assist certain populations to install and operate the apps, which could increase their uptake.

Mohr et al. [86] theorises some of these inconveniences could be solved if technologies are integrated into existing healthcare delivery systems. For example, by integrating them with electronic medical records, which would favour to use the resulting data for the patient's overall treatment.

3.7 User awareness, coherent emotional behaviour, empathy and trust

Recent reviews of conversational agents for health and well-being [76, 32], indicate the importance of establishing a **therapeutic alliance** with users to increase their satisfaction towards the system. Although this aspect is still not fully understood outside the human-human scenario, there exists an initial body of evidence that such alliance may be favoured by the inclusion of empathic reactions and the adaptation of the agent's behaviour to the users.

Early work in [87] emphasises the importance that the agents show social behaviours in order to from a strong therapeutic alliance, including the verbal (talk about the user-agent relationship, humour, greetings, smalltalk...) and nonverbal behaviours (direct body and facial orientation, gaze, smiles...).

Certainly, according to [88], a key characteristic of mental health agents is that they generate adequate emotional responses that convey **empathy** during the interaction with the user. Paiva et al. [89] present a very comprehensive study of empathy in virtual agents and robots, which covers different empathy mechanisms, modulation and expression.

Paiva et al. explore two ways for empathy: the agent being the target of empathy, and the agent showing empathy towards the user. Both aspects are interesting for mental health applications, in the first case, the agent evokes empathy in the user, which may be interesting for training social skills. In the latter, the agent responds congruently to the user emotional situation, which is key in order to develop adequate counsellors and coaches. As explained by the authors, empathy is very tightly coupled with user awareness, as the agent must identify the user's emotions and adapt its own affective response appropriately.

Martínez-Miranda et al. [88] distinguish between "therapeutic" and "natural" empathic agent reactions. For example, in the treatment of major depression agents should not show empathy by adopting a negative mood. Although it would be a sympathetic behaviour, it could be interpreted by the user as a agreement with their negative view of the world or the future. On the other hand, a "therapeutic empathy" would resemble a human therapist perspective when they incorporate not only emotional involvement but also a requited emotional detachment. In order to do so, the authors present a re-appraisal mechanism that includes projected emotional states and copying strategies.

An interesting means of showing empathy is to use humour. Ramakrishna et al. present a computational model of conversational humour in psychotherapy [90]. They consider different humour features including structural (words, word length...), stylistic (rhymes) and ambiguity that they included into the agent's prompts during a motivational interview. Their work was based on finding the best learning algorithm and does not report acceptability results with human users.

To create an alliance the system must encompass a mechanism of **user awareness** to adapt its behaviour and recommendations to the users and explain why they are meaningful to them. Abdulrahman and Richards [68] present a framework to build explainable agents using the FAtiMA architecture for affective agents. Their proposal includes a user model that considers the user's belief, goals, preferences, medical history and family context. Their system contains an "explanation engine" that decides when to deliver information and how to vary the explanation patterns according to the user model.

Bickmore et al. [91] also present a counselling framework based on an ontology that represents the user mental states and establishes the system actions as triggers that can modify them.

Other authors emphasise as well the need to consider patient's perspectives and needs to foster acceptability and uptake. Recently [92] presented a qualitative study with 29 participants with limited experience with chatbots and their role in health. The participants showed uncertainty about the trustworthiness and accuracy of the technology. The study covered their opinions about awareness, experience, perceived accuracy, maturity, security, anonymity, convenience and sign-posting. The authors conclude that for users to be receptive to this technology, user-centered approaches must be devised to address user concerns and engage patients with their health.

User awareness in the context of mental health must go a step further from general user-centered approaches to include an understanding of relevant factors that affect this population in particular. Burr and Morley [93] present a very detailed discussion on the use of the term "empower" with conversational agents that aim at "empowering" users to actively improve their mental health through a process of self-reflection (e.g. with the symptom monitoring, coaching or interviewing agents we have presented in Section 2). The authors claim that this presumes that these users want or even feel able to engage on such process. This may not be the case with users with mental health disorders and the authors propose to attend to important psycho-social factors that could have a negative effect on the ability of the user to engage in these interactions.

3.8 Scripted vs. data-based approaches

Rule-based conversational agents decide their responses based in if-then rules considering the users' inputs, while data-based approaches allow to learn the system responses from corpora and decide the next system intervention based on the most similar observed cases.

Abd-alrazaq et al. [31] found that most chatbots in their survey (>90%) were rule-based. The authors conclude that chatbots in mental health "lag behind chatbots in other fields" because the latter are using data-based statistical approaches. The authors argue that rule-based dialogues can be considered more secure but are also more restricted to the previously determined scenarios and have difficulties to adapt to new situations.

However, rule-based conversational agents have the advantage of providing predictable responses, which may be necessary in some mental health scenarios. In the case of interviewing agents, rule-based approaches make it possible

to produce heavily scripted and thus more replicable interviews , making sure that the agent has always followed the desired procedure without variations [50]. This allows to obtain better informed comparisons between interviews (of the same user at different times or between users).

Apart from this, an important factor that may lead to a modest presence of data-based approaches to build conversational agents for mental health is the small number of conversation datasets available.

[54] present the *Distress Analysis Interview Corpus* (DAIC), which contains annotated audio video recordings for clinical interviews to diagnose pscyhological distress. The corpus contains three interview types: face-to-face, wizard-of-oz and automated.

Morris et al. present the Koko platform and corpus [94], which promotes emotional resilience by passing messages between users who seek help and other users who want to give help. It has an automatic component that matches each incoming user input with similar posts in their existing corpus, if there is a highly rated match, they return the corresponding feedback to the user.

3.9 Standardized dialogue moves and reporting

The dialogue management component of conversational agents is created on the basis of a repertoir of system and user dialogue acts (abstract representation of the utterances). In the mental health area, they are usually created ad-hoc by the authors for their specific tasks. In the previous sections, we have surveyed different areas in which conversational agents can be employed. It would thus be interesting to have some common repertoires that would allow a better integration and comparison of systems as well as to reuse previous work by different authors.

Schulman et al. [95] present a conversational coach to motivate users to change to a more healthy behaviour. In their work, the authors adopt a *motivational interviewing* approach, in which the interviewer performs a client-centred interview that allows the user to think about their situation without offering explicit advice. A very interesting contribution is the dialogue acts taxonomy employed, which they used in the DTask dialogue manager. The taxonomy includes different types of system acts related to motivational interviewing, including evocative question, elaboration request, reflection, acknowledge of importance and summarization. For the user dialogue acts, they consider only statements, which may vary in valence (change or resistance, category (status-quo, change implication, change outlook and change intention) and content (the particular topic, concern or value). This is a very promising piece of work that could be reused in other agents with similar purposes. However, the authors state that it is only enough for a fragment of the meaning a trained counsellor may take from client utterances.

Lisetti et al. [96] address *brief interventions*: well-structured counselling sessions focused on a specific aspect of an unhealthy habit, in their case alcoholism. The authors develop a reinforcement learning dialogue manager structured in several stages. Their dialogue states have several features including whether the user has been greeted, which questions have been already posed (from a predefined interview), confidence level in the recognised input, etc. Their goal is to obtain an alcohol screening brief intervention based on their list of questions without a fixed dialogue plan, which can be updated dynamically depending on whether they find that the user has a dangerous drinking pattern or not.

Meguro et al. [97] present an active listening agent that satisfies the user desire to be listened to. In the mental health area, such system could help users to outsource their negative thoughts and explain themselves between feeling judged. In an attempt to comply with an standard, the dialogue acts employed are based on the DAMSL tag set. They are: self-disclosure, information, acknowledgement, question, sympathy and greeting. The authors present a very illustrative comparison between the frequency of these dialogue acts in casual vs. listening-oriented dialogues and found that the rates of *self-disclosure* and *information* were lower for the listening-oriented dialogues while *acknowledgement* and *question* were higher.

Bickmore et al. [91] define *TherapeuticDialiogueActions* (e.g. negotiate a specific goal for the near future) and *NonTehrapeuticDialogueActions* (e.g. greetings, small talk...). The counselling dialogues they establish follow a predefined high-level structure of several steps including: opening, social dialogue, review of previously assigned tasks and goals, assessing the user's state, counselling based on motivational interviewing, negotiating of new goals and tasks, closing and farewell.

The DAIC corpus mentioned previosly [54] includes dialogue annotation. It considers information about the points in the conversation when it was appropriate for the agent to provide feedback, as well as domain-specific dialogue acts to support follow-up, although they do not indicate the repertoire in the paper.

With respect to the standards in reporting, Vayidyam et al. [32] state the necessity of following standards in the studies that report the use of conversational systems for mental health, so that they can be conveniently compared, studied

and replicated. They evaluate the use of the World Health Organization mHealth Evidence Reporting and Assessment framework [98], but unfortunately it lacks specific aspects for conversational agents.

3.10 Ethical issues

Kretzschmar et al. [33] present a discussion on the ethical concerns that arise from the use of conversational agents in mental health support, and the perspectives of young persons from the Oxford Neuroscience, Ethics and Society Young People's Advisory Group. They reached to the conclusions the minimal ethical standards should include:

- Privacy and confidentiality. Keep personal information confidential, conversations de-identified, and privacy arrangements made transparent.
- Efficacy. The support should be based on scientific evidence, and such evidence should be available to users, who must be informed about what to expect from the conversational agent.
- Safety. User must know at all times that they are speaking with a machine and be encouraged to seek human support. The agent should include mechanisms to deal with emergency situations (e.g. risk of self-harm).

Privacy is one of the key issues for mental health conversational agents. Even when data is de-identified, when the systems are responsive to speech input, the user voice must be recorded and processed. Even if it is anonymised, a person may be recognised by their voice. In the speech community there is increasing interest on how to be able to deal with voice recordings, and share them (e.g. to create and share corpora to train better systems) at the same time as user anonymity is granted [99].

Stiefel's recent study on mental health confidentialy with chatbots [100] reveals that the current framework in the U.S. does not provide any obligation to these apps with regards to disclosure despite the restrictions in this matter that apply for licensed mental health professionals. Thus, additional confidentiality restrictions should be studied and posed. Miner et al. [73] also illustrate the case of disclosures that in some psychotherapy contexts make clinicians liable to civil judgement (e.g. non sharing homicidal ideation with the intended victim) and it is not clear how they apply to conversational agents.

Martínez-Martín and Kreitmair [101] present a comprehensive overview of the ethical issues for digital psychotherapy apps. One relevant aspect they cover is the efficacy and safety of the psychotherapy delivered, as incorrect advice may cause direct harm. They also mention the effect of a possible "commercialisation gap" in which the technology developed by clinicians is subject to more rigorous safety tests, while the ones developed by the private sector are designed to maximise user engagement and thus become more popular among users.

Vaidyam et al. [32] also discuss an interesting aspect which is many times neglected: the potential for users to become over-attached or dependent of the conversational agent, which may even be preventing them to establish face-face interactions with other persons. Hudlicka [9] also poses relevant questions about the relationships that may be maintained between the agents and their users, and whether they can lead to a false sense of trustworthiness, to a replacement of human relationships, or even to establishing a strong bond with the agent, which is not capable of real emotions.

Mulvenna et al. [102] present a manifesto with 12 principles as a starting point for ethical by design development of chatbots for mental health, which they have used to develop the iHelpr system [103]. These include among others empathy for users, providing informed decisions, respect to choose ways to be engaged, privacy and security, equitable access and complementary viewpoints, challenge possible biased incorporated into the system, support through lifespan and progression policy, planning for failure, transparency and reporting.

4 Conclusions

Conversational agents have demonstrated a great potential as to engage users in meaningful conversations in different areas. In the case of mental health, they can help individuals to access mental health services where they are difficult to access, reduce costs for continued interactions and monitoring, offering very valuable data for clinicians, and establish a setting that users may feel as not stigmatising and judgemental, thus making it more attractive specially for young users who find it difficult to disclose in person.

This technology can encourage users to take action an improve their mental health and well-being in a myriad of ways. In this chapter we have presented their main applications including professional training, fostering mental health literacy, symptom identification and monitoring, diagnosis, self-management of emotion, mood and mental health conditions, computer-aided psychotherapy, and behavioural interventions, counselling and coaching.

However, despite their benefits and multiple applications, the development of conversational agents for mental health poses numerous challenges, we have identified some of them and discussed the main approaches suggested in the literature and the avenues for future research. We have placed special emphasis on ethical issues, which are of paramount importance in this context, in order to generate systems that maintain privacy, do not raise false expectations, are accurate and based on scientific evidence and do not prevent people from reaching other possibly more effective mental health services.

In summary, conversational systems for mental health is a thrilling research area with many open questions and an enormous potential to do social good.

Acknowledgements

This research has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823907
(MENHIR project `https://menhir-project.eu`).

References

[1] Michael F. McTear, Zoraida Callejas, and David Griol. *The Conversational Interface. Talking to smart devices*. Springer, 2016.

[2] D. Griol, Zoraida Callejas, R. López-Cózar, and G. Riccardi. A domain-independent statistical methodology for dialog management in spoken dialog systems. *Computer Speech & Language*, 28(3):743–768, 2014.

[3] R. Pieraccini. *The Voice in the Machine: Building Computers that Understand Speech*. The MIT Press, 2012.

[4] Cathy Pearl. *Designing Voice User Interfaces: Principles of Conversational Experiences*. O'Reilly Media, 2017.

[5] Joao Luis Zeni Montenegro, Cristiano André da Costa, and Rodrigo da Rosa Righi. Survey of conversational agents in health. *Expert Systems with Applications*, 129:56–67, 2019.

[6] T. Bickmore and T. Giorgino. Health dialog systems for patients and consumers. *Journal of Biomedical Informatics*, 39(5):556–571, 2006.

[7] L. Laranjo, A.G. Dunn, H.L. Tong, A.B. Kocaballi, J. Chen, R. Bashir, D. Surian, B. Gallego, F. Magrabi, A.Y.S. Lau, and E. Coiera. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25(9):1248–1258, 2018.

[8] L. Pfeifer and T. Bickmore. Designing Embodied Conversational Agents to Conduct Longitudinal Health Interviews. In *Proc. of IVA'10*, pages 4698–4703, Philadelphia, USA, 2010.

[9] Eva Hudlicka. Virtual training and coaching of health behavior: Example from mindfulness meditation training. *Patient Education and Counseling*, 92(2):160–166, 2013.

[10] T. Giorgino, I. Azzini, C. Rognoni, S. Quaglini, M. Stefanelli, R. Gretter, and D. Falavigna. Automated spoken dialogue system for hypertensive patient home management. *Journal of Medical Informatics*, 74:159–167, 2004.

[11] K.H. Mooney, S.L. Beck, W.N. Dudley, R. Farzanfar, and R. Friedman. A computer-based telecommunication system to improve symptom care for women with breast cancer. *Annals of Behavioral Medicine Annual Meeting Supplement*, 27:152–161, 2004.

[12] J. Allen, G. Ferguson, N. Blaylock, D. Byron, N. Chambers, M. Dzikovska, L. Galescu, and M. Swift. Chester: towards a personal medication advisor. *Journal of Biomedical Informatics*, 39(5):500–513, 2006.

[13] Timothy W. Bickmore, Kathryn Puskar, Elizabeth A. Schlenk, Laura M. Pfeifer, and Susan M. Sereika. Maintaining reality: Relational agents for antipsychotic medication adherence. *Interacting with Computers*, 22(4):276–288, July 2010.

[14] A. Shamekhi, H. Trinh, T.W. Bickmore, T.R. Deangelis, T. Ellis, B.V. Houlihan, and N.K. Latham. A virtual self-care coach for individuals with spinal cord injury. In *Proc. of ASSETS'16*, pages 327–328, 2016.

[15] D. Griol and Z. Callejas. Mobile Conversational Agents for Context-Aware Care Applications. *Cognitive Computation*, 8(2):336–356, 2016.

[16] M.A. Sillice, P.J. Morokoff, G. Ferszt, T. Bickmore, B.C. Bock, R. Lantini, and W.F. Velicer. Using relational agents to promote exercise and sun protection: Assessment of participants' experiences with two interventions. *Journal of Medical Internet Research*, 20(2), 2018.

[17] R. Farzanfar, S. Frishkopf, J. Migneault, and R. Friedman. Telephone-linked care for physical activity: A qualitative evaluation of the use patterns of an information technology program for patients. *Journal of Biomedical Informatics*, 38:220–228, 2005.

[18] Antonio Benítez-Guijarro, Ángel Ruiz-Zafra, Zoraida Callejas, Nuria Medina-Medina, Kawtar Benghazi Akhlaki, and Manuel Noguera. General architecture for development of virtual coaches for healthy habits monitoring and encouragement. *Sensors*, 19(1):108, 2019.

[19] K.G. Ghanem, H. Hutton, J. Zenilman, R. Zimba, and E. Erbelding. Audio computer assisted self interview and face to face interview modes in assessing response bias among STD clinic patients. *Sexually Transmitted Infections*, 81(5):421–425, 2005.

[20] F. Ahmad, S. Hogg-Johnson, D. Stewart, H. Skinner, R. Glazier, and W. Levinson. Computer-assisted screening for intimate partner violence and control: a randomized trial. *Annals of Internal Medicine*, 151(2):93–102, 2009.

[21] T.W. Bickmore, S.E. Mitchell, B.W. Jack, M.K. Paasche-Orlow, L.M. Pfeifer, and J. O'Donnell. Response to a relational agent by hospital patients with depressive symptoms. *Interacting with Computers*, 22:289–298, 2010.

[22] K. López de Ipiña, J.B. Alonso, J. Solé-Casals, N. Barroso, P. Henriquez, M. Faundez-Zanuy, C.M. Travieso, M. Ecay-Torres, P. Martínez-Lage, and H. Eguiraun. On Automatic Diagnosis of Alzheimer's Disease Based on Spontaneous Speech Analysis and Emotional Temperature. *Cognitive Computation*, 7(1):44–55, 2015.

[23] S. Payr. Closing and closure in human-companion interactions: Analyzing video data from a field study. In *Proc. of RO-MAN'10*, pages 476–481, Viareggio, Italy, 2010.

[24] M. Cavazza, R. Santos de la Cámara, and M. Turunen. How Was Your Day? a Companion ECA. In *Proc. of AAMAS'10*, pages 1629–1630, Toronto, Canada, 2010.

[25] S. Young. Cognitive user interfaces. *IEEE Signal Processing Magazine*, 27(3):128–140, 2011.

[26] I. Leite, A. Pereira, G. Castellano, S. Mascarenhas, C. Martinho, and A. Paiva. Modelling empathy in social robotic companions. *Advances in User Modeling*, 7138:135–147, 2012.

[27] A. Sixsmith, S. Meuller, F. Lull, M. Klein, I. Bierhoff, S. Delaney, and R. Savage. SOPRANO - An Ambient Assisted Living System for Supporting Older People at Home. In *Proc. of ICOST'09*, pages 233–236, Tours, France, 2009.

[28] E. Andre, E. Bevacqua, D.K.J. Heylen, R. Niewiadomski, C. Pelachaud, C. Peters, I. Poggi, and M. Rehm. *Emotion Oriented Systems. The Humaine Handbook. Cognitive Technologies*, chapter Non-verbal Persuasion and Communication in an Affective Agent, pages 585–608. Springer Verlag, 2011.

[29] T. Rehrl, J. Geiger, M. Golcar, S. Gentsch, J. Knobloch, G. Rigoll, K. Scheibl, W. Schneider, S. Ihsen, and F. Wallhoff. The Robot ALIAS as a Database for Health Monitoring for Elderly People. In *Proc. of AAL'13*, pages 414–423, Berlin, Germany, 2013.

[30] Becky Inkster, Shubhankar Sarda, and Vinod Subramanian. 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):e12106, 2018.

[31] Alaa A. Abd-alrazaq, Mohannad Alajlani, Ali Abdallah Alalwan, Bridgette M. Bewick, Peter Gardner, and Mowafa Househ. An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics*, 132:103978, 2019.

[32] Aditya Nrusimha Vaidyam, Hannah Wisniewski, John David Halamka, Matcheri S. Kashavan, and John Blake Torous. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *The Canadian Journal of Psychiatry*, page 070674371982897, 2019.

[33] Kira Kretzschmar, Holly Tyroll, Gabriela Pavarini, Arianna Manzini, Ilina Singh, and NeurOx Young People's Advisory Group. Can Your Phone Be Your Therapist? Young People's Ethical Perspectives on the Use of Fully Automated Conversational Agents (Chatbots) in Mental Health Support. *Biomedical Informatics Insights*, 11:117822261982908, 2019.

[34] Simon Provoost, Ho Ming Lau, Jeroen Ruwaard, and Heleen Riper. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. *Journal of Medical Internet Research*, 19(5):151, 2017.

[35] 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. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25(9):1248–1258, 2018.

[36] Stephanie Carnell, Shivashankar Halan, Michael Crary, Aarthi Madhavan, and Benjamin Lok. Adapting Virtual Patient Interviews for Interviewing Skills Training of Novice Healthcare Students. In Willem-Paul Brinkman, Joost Broekens, and Dirk Heylen, editors, *Proc. of IVA'15*, pages 50–59. Springer International Publishing, 2015.

[37] Shivashankar Halan, Isaac Sia, Michael Crary, and Benjamin Lok. Exploring the Effects of Healthcare Students Creating Virtual Patients for Empathy Training. In Willem-Paul Brinkman, Joost Broekens, and Dirk Heylen, editors, *Proc. of IVA'15*, Lecture Notes in Computer Science, pages 239–249, Cham, 2015. Springer International Publishing.

[38] Andrew Cordar, Michael Borish, Adriana Foster, and Benjamin Lok. Building Virtual Humans with Back Stories: Training Interpersonal Communication Skills in Medical Students. In Timothy Bickmore, Stacy Marsella, and Candace Sidner, editors, *Intelligent Virtual Agents*, volume 8637, pages 144–153. Springer International Publishing, Cham, 2014.

[39] Cameron Carpenter, Leticia Osterberg, and Geoff Sutcliffe. SAMHT - Suicidal Avatars for Mental Health Training. In *Proc. of FLAIRS'12*, pages 484–487, Marco Island, Florida, USA, 2012. AAAI Publications.

[40] Nancy D. Berkman, Stacey L. Sheridan, Katrina E. Donahue, David J. Halpern, and Karen Crotty. Low Health Literacy and Health Outcomes: An Updated Systematic Review. *Annals of Internal Medicine*, 155(2):97, 2011.

[41] Stan Kutcher, Yifeng Wei, and Connie Coniglio. Mental Health Literacy: Past, Present, and Future. *The Canadian Journal of Psychiatry*, 61(3):154–158, 2016.

[42] Claire M Kelly, Anthony F Jorm, and Annemarie Wright. Improving mental health literacy as a strategy to facilitate early intervention for mental disorders. *Medical Journal of Australia*, 187(S7), 2007.

[43] Anthony F. Jorm. Why we need the concept of "mental health literacy". *Health Communication*, 30(12):1166–1168, 2015.

[44] A. F. Jorm. Mental health literacy: Public knowledge and beliefs about mental disorders. *British Journal of Psychiatry*, 177(5):396–401, 2000.

[45] Timothy W. Bickmore, Laura M. Pfeifer, and Michael K. Paasche-Orlow. Using computer agents to explain medical documents to patients with low health literacy. *Patient Education and Counseling*, 75(3):315–320, June 2009.

[46] Jing Ling Tay, Yi Fen Tay, and Piyanee Klainin-Yobas. Effectiveness of information and communication technologies interventions to increase mental health literacy: A systematic review. *Early Intervention in Psychiatry*, 12(6):1024–1037, 2018.

[47] Joel Sebastian and Deborah Richards. Changing stigmatizing attitudes to mental health via education and contact with embodied conversational agents. *Computers in Human Behavior*, 73, 2017.

[48] Yifeng Wei, Patrick J. McGrath, Jill Hayden, and Stan Kutcher. Mental health literacy measures evaluating knowledge, attitudes and help-seeking: a scoping review. *BMC Psychiatry*, 15(1):291, 2015.

[49] Saul Shiffman, Arthur A. Stone, and Michael R. Hufford. Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, 4(1):1–32, 2008.

[50] Pierre Philip, Jean-Arthur Micoulaud-Franchi, Patricia Sagaspe, Etienne De Sevin, Jérôme Olive, Stéphanie Bioulac, and Alain Sauteraud. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. *Scientific Reports*, 7:42656, 2017.

[51] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, Gale Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Scherer, Giota Stratou, Apar Suri, David Traum, Rachel Wood, Yuyu Xu, Albert Rizzo, and Louis-Philippe Morency. SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support. In *Proc. of AAMAS'14*, pages 1061–1068, Paris, France, 2014.

[52] Alexandros Roniotis and Manolis Tsiknakis. Detecting Depression Using Voice Signal Extracted by Chatbots: A Feasibility Study. In Anthony L. Brooks, Eva Brooks, and Nikolas Vidakis, editors, *Interactivity, Game Creation, Design, Learning, and Innovation*, volume 229, pages 386–392. Springer International Publishing, Cham, 2018.

[53] S.H. Kang and J. Gratch. Socially anxious people reveal more personal information with virtual counselors that talk about themselves using intimate human back stories. *Studies in Health Technology and Informatics*, 181:202–206, 2012.

[54] Jonathan Gratch, Ron Artstein, Gale Lucas, Giota Stratou, Stefan Scherer, Angela Nazarian, Rachel Wood, Jill Boberg, David DeVault, Stacy Marsella, David Traum, Skip Rizzo, and Louis-Philippe Morency. The Distress Analysis Interview Corpus of human and computer interviews. In *Proc. of LREC'14*, pages 3123–3128, Reykjavik, Iceland, 2014.

[55] Matthew D. Pickard, Catherine A. Roster, and Yixing Chen. Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? *Computers in Human Behavior*, 65:23–30, 2016.

[56] Gale M. Lucas, Albert Rizzo, Jonathan Gratch, Stefan Scherer, Giota Stratou, Jill Boberg, and Louis-Philippe Morency. Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers. *Frontiers in Robotics and AI*, 4:51, 2017.

[57] Albert Rizzo, Belinda Lange, John G. Buckwalter, Eric Forbell, Julia Kim, Kenji Sagae, Josh Williams, JoAnn Difede, Barbara O. Rothbaum, Greg Reger, Thomas Parsons, and Patrick Kenny. SimCoach: an intelligent virtual human system for providing healthcare information and support. *International Journal on Disability and Human Development*, 10(4), 2011.

[58] Juan Martínez-Miranda. Embodied Conversational Agents for the Detection and Prevention of Suicidal Behaviour: Current Applications and Open Challenges. *Journal of Medical Systems*, 41(9):135, 2017.

[59] MarySue V Heilemann, Adrienne Martinez, and Patricia D Soderlund. A Mental Health Storytelling Intervention Using Transmedia to Engage Latinas: Grounded Theory Analysis of Participants' Perceptions of the Story's Main Character. *Journal of Medical Internet Research*, 20(5):e10028, 2018.

[60] David Bakker and Nikki Rickard. Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. *Journal of Affective Disorders*, 227:432–442, 2018.

[61] Seung-A. Annie Jin. The effects of incorporating a virtual agent in a computer-aided test designed for stress management education: The mediating role of enjoyment. *Computers in Human Behavior*, 26(3):443–451, 2010.

[62] Melissa D. Pinto, Ronald L. Hickman, John Clochesy, and Marc Buchner. Avatar-based depression self-management technology: promising approach to improve depressive symptoms among young adults. *Applied Nursing Research*, 26(1):45–48, 2013.

[63] Lazlo Ring, Barbara Barry, Kathleen Totzke, and Timothy Bickmore. Addressing Loneliness and Isolation in Older Adults: Proactive Affective Agents Provide Better Support. In *Proc. of ACII'13*, pages 61–66, Geneva, Switzerland, 2013.

[64] J. Proudfoot, D. Goldberg, A. Mann, B. Everitt, I. Marks, and J. A. Gray. Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. *Psychological Medicine*, 33(2):217–227, 2003.

[65] Adam Miner, Amanda Chow, Sarah Adler, Ilia Zaitsev, Paul Tero, Alison Darcy, and Andreas Paepcke. Conversational Agents and Mental Health: Theory-Informed Assessment of Language and Affect. In *Proc. of HAI'16*, pages 123–130, Biopolis, Singapore, 2016. ACM Press.

[66] Tristan Gorrindo and James E. Groves. Computer Simulation and Virtual Reality in the Diagnosis and Treatment of Psychiatric Disorders. *Academic Psychiatry*, 33(5):413–417, 2009.

[67] Morie Jacquelyn Ford, Antonisse Jamie, Bouchard Sean, and Chance Eric. Virtual Worlds as a Healing Modality for Returning Soldiers and Veterans. *Studies in Health Technology and Informatics*, pages 273–276, 2009.

[68] Amal Abdulrahman and Deborah Richards. Modelling Therapeutic Alliance using a User-aware Explainable Embodied Conversational Agent to Promote Treatment Adherence. In *Proc. of IVA'19*, pages 248–251, Paris, France, 2019.

[69] T. Kuosmanen, T.M. Fleming, J. Newell, and M.M. Barry. A pilot evaluation of the sparx-r gaming intervention for preventing depression and improving wellbeing among adolescents in alternative education. *Internet Interventions*, 8:40 – 47, 2017.

[70] Christopher Burton, Aurora Szentagotai Tatar, Brian McKinstry, Colin Matheson, Silviu Matu, Ramona Moldovan, Michele Macnab, Elaine Farrow, Daniel David, Claudia Pagliari, Antoni Serrano Blanco, Maria Wolters, and for the Help4Mood Consortium. Pilot randomised controlled trial of Help4mood, an embodied virtual agent-based system to support treatment of depression. *Journal of Telemedicine and Telecare*, 22(6):348–355, 2016.

[71] Adam S. Miner, Arnold Milstein, Stephen Schueller, Roshini Hegde, Christina Mangurian, and Eleni Linos. Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health. *JAMA Internal Medicine*, 176(5):619, 2016.

[72] Jurek Kirakowski, Patrick O'Donnell, and Anthony Yiu. The Perception of Artificial Intelligence as "Human" by Computer Users. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, and Julie A. Jacko, editors, *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments*, volume 4552, pages 376–384. Berlin, Heidelberg, 2007.

[73] Adam S. Miner, Nigam Shah, Kim D. Bullock, Bruce A. Arnow, Jeremy Bailenson, and Jeff Hancock. Key Considerations for Incorporating Conversational AI in Psychotherapy. *Frontiers in Psychiatry*, 10:746, 2019.

[74] Ewa Luger and Abigail Sellen. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proc. of CHI '16*, pages 5286–5297, Santa Clara, California, USA, 2016.

[75] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. Patterns for How Users Overcome Obstacles in Voice User Interfaces. In *Proc. of CHI '18*, pages 1–7, Montreal QC, Canada, 2018.

[76] J. Kim, S. J. Park, and P. L. Robert. Conversational Agents for Health and Wellbeing: Review and Future Agendas. In *Proc. of CSCW'19*, Austin, Texas, USA, 2019.

[77] Amit Baumel, Amanda Tinkelman, Nandita Mathur, and John M Kane. Digital Peer-Support Platform (7cups) as an Adjunct Treatment for Women With Postpartum Depression: Feasibility, Acceptability, and Preliminary Efficacy Study. *JMIR mHealth and uHealth*, 6(2):e38, 2018.

[78] Eileen Bendig, Benjamin Erb, Lea Schulze-Thuesing, and Harald Baumeister. The Next Generation: Chatbots in Clinical Psychology and Psychotherapy to Foster Mental Health – A Scoping Review. *Verhaltenstherapie*, pages 1–13, 2019.

[79] Evonne Miller and Debra Polson. Apps, Avatars, and Robots: The Future of Mental Healthcare. *Issues in Mental Health Nursing*, 40(3):208–214, 2019.

[80] Timothy W. Bickmore and Rosalind W. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, 12(2):293–327, 2005.

[81] Maja J. Matarić and Brian Scassellati. Socially Assistive Robotics. In Bruno Siciliano and Oussama Khatib, editors, *Springer Handbook of Robotics*, pages 1973–1994. Springer International Publishing, Cham, 2016.

[82] Iolanda Leite, Carlos Martinho, and Ana Paiva. Social Robots for Long-Term Interaction: A Survey. *International Journal of Social Robotics*, 5(2):291–308, 2013.

[83] Ryan M. Schuetzler, G. Mark Grimes, and Justin Scott Giboney. The effect of conversational agent skill on user behavior during deception. *Computers in Human Behavior*, 97:250–259, 2019.

[84] Steven Chan, Luming Li, John Torous, David Gratzer, and Peter M. Yellowlees. Review of Use of Asynchronous Technologies Incorporated in Mental Health Care. *Current Psychiatry Reports*, 20(10):85, 2018.

[85] John Torous, Hannah Wisniewski, Gang Liu, and Matcheri Keshavan. Mental Health Mobile Phone App Usage, Concerns, and Benefits Among Psychiatric Outpatients: Comparative Survey Study. *JMIR Mental Health*, 5(4):e11715, 2018.

[86] David C. Mohr, Michelle Nicole Burns, Stephen M. Schueller, Gregory Clarke, and Michael Klinkman. Behavioral Intervention Technologies: Evidence review and recommendations for future research in mental health. *General Hospital Psychiatry*, 35(4):332–338, 2013.

[87] Timothy Bickmore and Amanda Gruber. Relational Agents in Clinical Psychiatry:. *Harvard Review of Psychiatry*, 18(2):119–130, 2010.

[88] Juan Martinez-Miranda, Adrian Breso, and Juan Miguel Garcia-Gomez. Look on the Bright Side: A Model of Cognitive Change in Virtual Agents. In *Proc. of IVA'14*, pages 285–294, Cham, Germany, 2014.

[89] Ana Paiva, Iolanda Leite, Hana Boukricha, and Ipke Wachsmuth. Empathy in Virtual Agents and Robots: A Survey. *ACM Transactions on Interactive Intelligent Systems*, 7(3):1–40, 2017.

[90] Anil Ramakrishna, Timothy Greer, David Atkins, and Shrikanth Narayanan. Computational Modeling of Conversational Humor in Psychotherapy. In *Proc. of Interspeech'18*, pages 2344–2348, 2018.

[91] Timothy W. Bickmore, Daniel Schulman, and Candace L. Sidner. A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. *Journal of Biomedical Informatics*, 44(2):183–197, 2011.

[92] Tom Nadarzynski, Oliver Miles, Aimee Cowie, and Damien Ridge. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *Digital Health*, 5:205520761987180, January 2019.

[93] Christopher Burr and Jessica Morley. Empowerment or Engagement? Digital Health Technologies for Mental Healthcare. *SSRN Electronic Journal*, 2019.

[94] Robert R Morris, Kareem Kouddous, Rohan Kshirsagar, and Stephen M Schueller. Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions. *Journal of Medical Internet Research*, 20(6):e10148, 2018.

[95] Daniel Schulman, Timothy Bickmore, and Candace Sidner. An intelligent conversational agent for promoting long-term health behavior change using motivational interviewing. In *AAAI Spring Symposium Series*, 2011.

[96] Christine Lisetti, Reza Amini, Ugan Yasavur, and Naphtali Rishe. I Can Help You Change! An Empathic Virtual Agent Delivers Behavior Change Health Interventions. *ACM Transactions on Management Information Systems*, 4(4):1–28, 2013.

[97] Toyomi Meguro, Ryuichiro Higashinaka, Kohji Dohsaka, Yasuhiro Minami, and Hideki Isozaki. Analysis of Listening-oriented Dialogue for Building Listening Agents. In *Proc. of SIGDIAL'09 Conference*, pages 124–127, London, UK, 2009.

[98] Smisha Agarwal, Amnesty E LeFevre, Jaime Lee, Kelly L'Engle, Garrett Mehl, Chaitali Sinha, and Alain Labrique. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. *BMJ*, page i1174, 2016.

[99] Andreas Nautsch, Catherine Jasserand, Els Kindt, Massimiliano Todisco, Isabel Trancoso, and Nicholas Evans. The GDPR and Speech Data: Reflections of Legal and Technology Communities, First Steps Towards a Common Understanding. In *Proc. of Interspeech'19*, pages 3695–3699, 2019.

[100] Scott Stiefel. 'The Chatbot Will See You Now': Mental Health Confidentiality Concerns in Software Therapy. *SSRN Electronic Journal*, 2018.

[101] Nicole Martinez-Martin and Karola Kreitmair. Ethical Issues for Direct-to-Consumer Digital Psychotherapy Apps: Addressing Accountability, Data Protection, and Consent. *JMIR Mental Health*, 5(2):e32, 2018.

[102] Maurice Mulvenna, Jennifer Boger, and Raymond Bond. Ethical by Design: A Manifesto. In *Proc. of ECCE'17*, pages 51–54, Umeå, Sweden, 2017.

[103] Gillian Cameron, David Cameron, Gavin Megaw, Raymond Bond, Maurice Mulvenna, Siobhan O'Neill, Cherie Armour, and Michael McTear. Best Practices for Designing Chatbots in Mental Healthcare - A Case Study on iHelpr. 2018.