


Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape

The Canadian Journal of Psychiatry /
La Revue Canadienne de Psychiatrie
2019, Vol. 64(7) 456-464
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0706743719828977
TheCJP.ca | LaRCP.ca



« Chatbots » et agents conversationnels en santé mentale : une revue de l'environnement psychiatrique

Aditya Nrusimha Vaidyam, BS¹ , Hannah Wisniewski, BS¹,
John David Halamka, MD¹, Matcheri S. Kashavan¹,
and John Blake Torous, MD, MBI¹

Abstract

Objective: The aim of this review was to explore the current evidence for conversational agents or chatbots in the field of psychiatry and their role in screening, diagnosis, and treatment of mental illnesses.

Methods: A systematic literature search in June 2018 was conducted in PubMed, EmBase, PsycINFO, Cochrane, Web of Science, and IEEE Xplore. Studies were included that involved a chatbot in a mental health setting focusing on populations with or at high risk of developing depression, anxiety, schizophrenia, bipolar, and substance abuse disorders.

Results: From the selected databases, 1466 records were retrieved and 8 studies met the inclusion criteria. Two additional studies were included from reference list screening for a total of 10 included studies. Overall, potential for conversational agents in psychiatric use was reported to be high across all studies. In particular, conversational agents showed potential for benefit in psychoeducation and self-adherence. In addition, satisfaction rating of chatbots was high across all studies, suggesting that they would be an effective and enjoyable tool in psychiatric treatment.

Conclusion: Preliminary evidence for psychiatric use of chatbots is favourable. However, given the heterogeneity of the reviewed studies, further research with standardized outcomes reporting is required to more thoroughly examine the effectiveness of conversational agents. Regardless, early evidence shows that with the proper approach and research, the mental health field could use conversational agents in psychiatric treatment.

Abrégé

Objectif : Cette revue visait à explorer les données probantes actuelles sur les agents conversationnels ou les « chatbots » (robots parleurs) dans le domaine de la psychiatrie et le rôle que jouent ceux-ci dans le dépistage, le diagnostic, et le traitement des maladies mentales.

Méthode : Une recherche systématique de la littérature a été menée en juin 2018 dans PubMed, EmBase, PsycINFO, Cochrane, Web of Science, et IEEE Xplore. Les études incluses portaient sur un « chatbot » dans un milieu de santé mentale axé sur les populations souffrant de dépression, d'anxiété, de schizophrénie, du trouble bipolaire et des troubles d'abus de substances ou qui étaient à risque élevé de développer un de ces troubles.

Résultats : Dans les bases de données choisies, 1466 dossiers ont été extraits et 8 études satisfaisaient aux critères d'inclusion. Deux études additionnelles ont été ajoutées après une sélection dans la liste de références, pour un total de 10 études incluses. En général, le potentiel de l'utilisation d'agents conversationnels en psychiatrie était estimé élevé dans toutes

¹ Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

Corresponding Author:

John Torous, MD, MBI, Beth Israel Deaconess Medical Center, Harvard Medical School, 75 Fenwood Road, Boston, MA, USA.
Email: jtorous@bidmc.harvard.edu

les études. En particulier, les agents conversationnels indiquaient le potentiel de se révéler bénéfiques en psychoéducation et en auto-engagement. En outre, le taux de satisfaction à l'égard des « chatbots » était élevé dans toutes les études, ce qui suggère qu'ils constitueraient un outil efficace et agréable dans un traitement psychiatrique.

Conclusion : Les données probantes préliminaires de l'utilisation psychiatrique des « chatbots » sont favorables. Cependant, étant donné l'hétérogénéité des études examinées, il faut plus de recherche contenant des résultats normalisés afin d'examiner minutieusement l'efficacité des agents conversationnels. Néanmoins, les premières données probantes indiquent qu'avec l'approche appropriée et la recherche, le domaine de la santé mentale pourrait utiliser les agents conversationnels dans le traitement psychiatrique.

Keywords

chatbot, conversational agent, embodied conversational agent, mental health, psychiatry, depression, medical informatics

Introduction

Access to mental health services and treatment remains an issue in all countries and cultures across the globe. Worldwide, major depression is the leading cause of years lived with disability and the fourth leading cause of disability-adjusted life years (DALYs).¹ According to the Health Canada Editorial Board on Mental Illnesses in Canada, more than 20% of Canadians will suffer from a mental illness during their lifetime, and the global economic burden of mental health in 2010 was estimated at 2.5 trillion US dollars.² With rates of suicide now actually increasing in many countries like the United States,³ it is clear that there is a need for new solutions and innovation in mental health.

Unfortunately, the current clinical workforce is insufficient in meeting these needs. There are approximately 9 psychiatrists per 100,000 people in developed countries¹ and as few as 0.1 for every 1,000,000⁴ in lower-income countries. This inadequacy in meeting the present or future demand for care has led to the proposal of technology as a solution. Particularly, there is a growing interest surrounding chatbots, also known as conversational agents or multipurpose virtual assistants.

Chatbots or conversational agents are here defined as digital tools existing either as hardware (such as an Amazon Echo running the Alexa digital assistant software) or software (such as Google Assistant running on Android devices or Siri running on Apple devices) that use machine learning and artificial intelligence methods to mimic humanlike behaviours and provide a task-oriented framework with evolving dialogue able to participate in conversation (see Figure 1). Gaining traction in the popular press, their potential in mental health is well represented, considering one of the top requests to Alexa during the summer of 2017 was “Alexa, help me relax,” according to the *MIT Technology Review*.⁵

This human-computer interaction technology was established academically half a century ago. In 1964, the programmable natural language processing program ELIZA was developed at the MIT Artificial Intelligence laboratory by Joseph Weizenbaum. Designed to act as a Rogerian psychotherapist, ELIZA could not understand the content of its conversations. However, many who used this chatbot believed it to be intelligent enough to comprehend

conversation and even became emotionally attached to it. Weizenbaum would later remark that “[he] had not realized . . . that extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people.”⁶ In 1972 at Stanford University, psychiatrist Kenneth Colby developed PARRY, a program capable of simulating the behaviour of a human with schizophrenia that was then “counseled” several times by ELIZA.

Fifty years later, the technology that made ELIZA possible is now available on the smartphones and smart home devices owned by billions around the world. According to market research, over three-quarters of Canadians own a smartphone, and already nearly 10% own a smart home device, such as the Google Home or Amazon Echo.⁷ The technology has also advanced to the point that chatbots today incorporate natural language processing for speech, removing the need for a keyboard, as anyone who has ever used Siri can affirm.

Although there is still much to be explored when it comes to chatbots in mental health, their potential has already begun to surface. Chatbots are being used in suicide prevention⁸ and cognitive-behavioural therapy,⁹ and they are even being tailored to certain populations, such as HARR-E and Wysa.¹⁰ In particular, chatbots may be helpful in providing treatment for those who are uncomfortable disclosing their feelings to a human being. Therefore, virtual therapy provided by a chatbot could not only improve access to mental health treatment but also be more effective for those reluctant to speak with a therapist. Veterans, for example, who are often reluctant to open up after a tour of duty, were significantly more likely to open up to a chatbot when told it was a virtual therapist than those who were told the chatbot was being controlled by a person¹¹—offering the potential to increase needed access to care.¹²

With increased access to technology and the ease of use that accompanies, interest in mental health chatbots has reached a point where some have labelled them “the future of therapy.”¹³ However, there is no consensus on the definition of psychiatric chatbots or their role in the clinic. While they do hold potential, little is known about who actually uses them and what their therapeutic effect may be. Evaluation efforts are further complicated by the rapid pace of

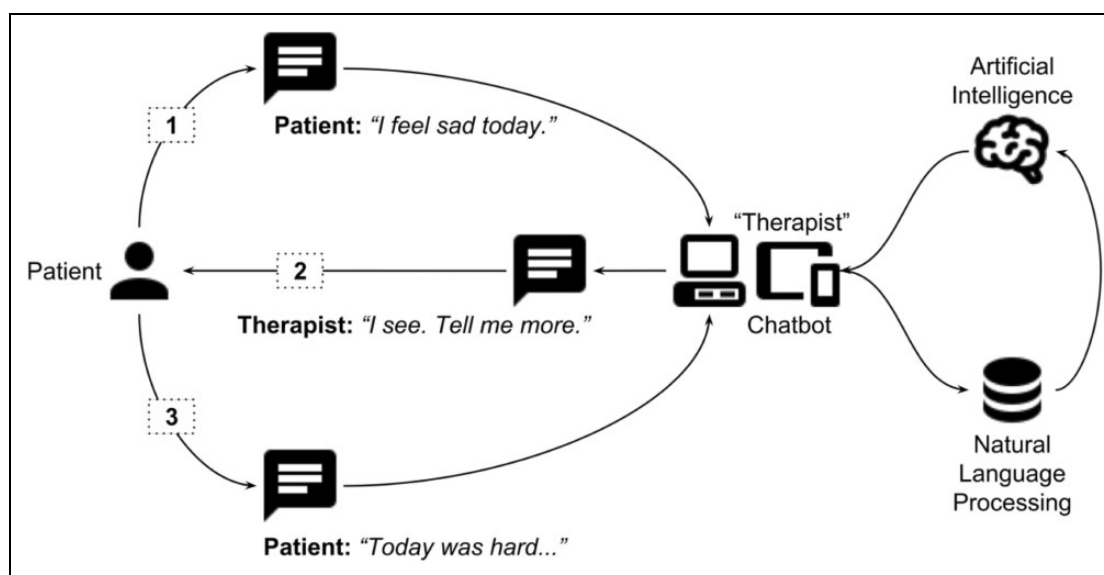


Figure 1. A sample interaction between a patient and a chatbot therapist.

development in hardware and that such software may behave and respond differently depending on region. For example, when a user said he or she felt sad, one chatbot, the US-developed Google Assistant, replied, "I wish I had arms so I could give you a hug," where the Russian-developed chatbot Alisa replied with "No one said life was about having fun."¹⁴ In this review, we explore today's early evidence for chatbots in mental health and focus on fundamental issues that need to be addressed for further research.

Methods

A librarian at the Boston University School of Medicine assisted in generation of a search term for selected databases (PubMed, EmBase, PsycINFO, Cochrane, Web of Science, and IEEE Xplore) using a combination of keywords, including *conversational agent* or *chatbot*, without other search restrictions or applied filters. The search for English-language papers, excluding conference abstracts, was conducted on July 11, 2018, and inclusion criteria for studies were those who used chatbots offering primarily dynamic user input and output response based on written, spoken, or animation of a 3D avatar ("embodiment") toward the care of those with or at high risk for depression, anxiety, schizophrenia, bipolar, and substance abuse disorders. Chatbots whose dialogue was prewritten but assembled and matched to the user input in a dynamic manner qualified, while those whose dialogue was recycled and derived from a human operator or peer (called the "Wizard of Oz" method) did not. Given the heterogeneity of chatbots, diversity of mental health applications, and variety of study methodologies in this nascent field, we aimed to select and review papers most relevant to clinical use-cases but acknowledge that firm boundaries are difficult to draw. Abstract and full-text screening, as well as data extraction phases, was conducted

by the first and last authors of this review independently and any disagreements resolved through discussion and consensus.

Results

Of the 1466 references identified from search terms applied to selected databases, 1066 conference proceedings and 104 duplicates were removed, leaving a total of 296 studies for the abstract screening phase. Through criteria-based abstract screening, 10 studies were identified for full-text screening, and 2 studies were selected postscreening from searched reference lists.¹⁵⁻²⁴ Ten studies were identified as relevant and entered the data extraction phase. The detailed PRISMA diagram in Figure 2 further outlines number of studies excluded per criteria. Listed for each selected study are summary data in Table 1, measures used in Table 2, application designations in Table 3, and functions or roles in Table 4. As only 2 of 10 selected studies provided full education and ethnicity demographic information, this information could not be summarized.

In contrast to the vast potential in marketing of chatbots to consumers, we found the academic psychiatric literature to be surprisingly sparse. The majority of research on chatbots appears to be conducted outside of traditional medical publication outlets, as we excluded nearly 75% of studies from our initial search that were largely engineering conference proceedings. The fact that much of this research is happening in other disciplines highlights a need to continue to bridge the multiple stakeholders advancing this work and create opportunities in this space for synergy. Still, from the 10 studies we identified, it is possible to comment on potential benefits as well as harms, new frontiers, ethical implications, and current limitations for chatbots in mental health.

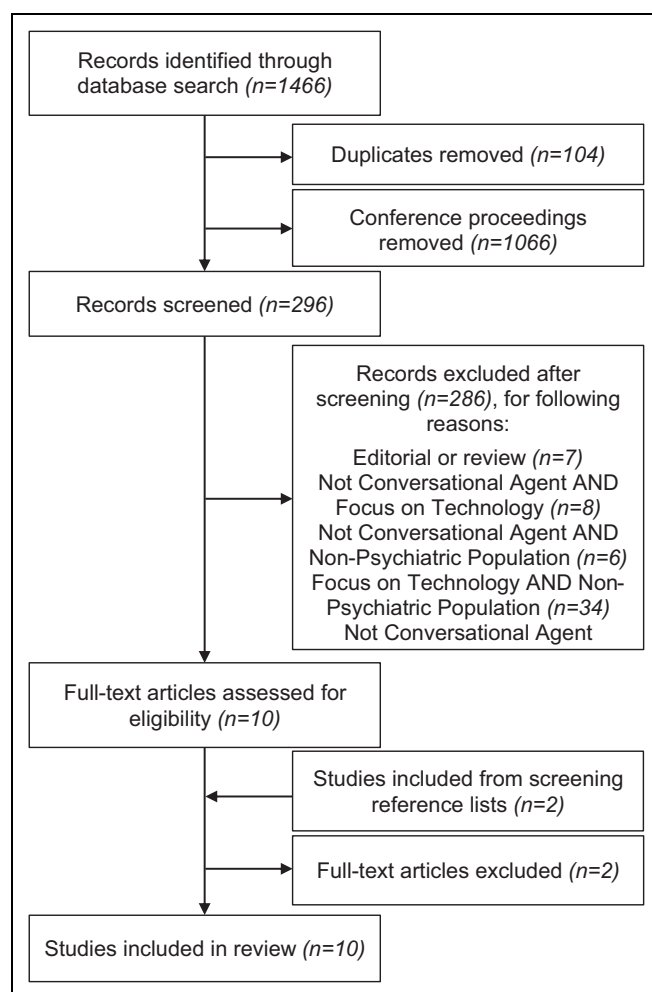


Figure 2. PRISMA diagram.

One potential benefit of conversational agents that was demonstrated includes self-psychoeducation and adherence. This included a variety of methods, such as tracking medication and physical activity adherence,²² providing cognitive behavioural therapy (CBT),¹⁵ and delivering healthy lifestyle recommendations¹⁹ across clinical and nonclinical groups. For example, Ly et al.¹⁵ found that a higher adherence to the conversational agent in a nonclinical population showed a significant improvement in psychological well-being and perceived stressed compared to those who did not use the intervention. In addition, Bickmore et al.²² found that individuals with major depressive disorder rated the therapeutic alliance between a conversational agent significantly higher than a clinician.

Furthermore, the participants in these studies showed high satisfaction with the interventions they received. Participants reported the interventions as helpful, easy to use, and informative¹⁹ and rated satisfaction highly (>4.2 out of 5) on all scales, including ease of use, desire to continue using the system, liking, and trust.²¹

A few studies examined the effectiveness of conversational agents in the diagnosis and treatment of psychiatric

disorders. Not only were conversational agents found to significantly reduce depressive symptoms in individuals with major depressive disorder,¹⁸ but an embodied conversational agent was also found to be able to efficiently identify patients with depressive symptoms.²⁴ Participants in both of these studies reported that the conversational agents were helpful and useable.

Our results indicate that the risk of harm from the use of chatbots is extremely low, with a total adverse event incidence of 1 in 759 recruited participants. The single adverse event, reported in Bickmore et al.,²¹ involved a participant who developed paranoia and withdrew from study. Another participant in the same study almost withdrew due to concerns of personal data theft until reorientation and counsel from the nurse on-call, suggesting a possible benefit to available clinician support.

Discussion

The results of these studies show that there is potential for effective, enjoyable mental health care using chatbots. However, the high heterogeneity in both the results and methodologies of these studies suggests further research is required to fully understand the best methods for implementation of a chatbot. Regardless, we were able to identify common benefits and potential harms of chatbot use.

Benefits

Overall, satisfaction with and potential for psychiatric use was reported to be high across all studies. Two of the common benefits of chatbots use were psychoeducation and adherence. Although these studies examine these factors by different modalities, they demonstrate that chatbots have the potential for individuals to provide self-care in both clinical and nonclinical populations, potentially alleviating the insufficiency of the workforce as previously mentioned. Furthermore, the positive user satisfaction results demonstrate not only that conversational agents have the potential to be used for self-adherence and education but also that users of these systems would find benefit in and enjoy doing so. Most important, the effectiveness of chatbots with individuals with major depressive disorder further suggests that chatbots would be feasible to use in clinical populations. Although the nature of these studies is very heterogeneous, the positive results from each provide evidence that conversational agents show promise in the psychiatry field.

Another advantage of chatbots for psychiatric applications is that they may be able to offer services to those who would not otherwise seek care because of stigma or cost.²¹ Although we found no studies that verifiably measured the effect on patient interactions with a chatbot, the anonymity offered by chatbots led some patients to disclose more sensitive information than they would to a human therapist, as demonstrated by Lucas et al.²³ in their study concerning reporting of posttraumatic stress disorder symptoms to a

Table 1. Reported Information about Each Selected Study and Chatbot Where Provided.

Study	Duration	Chatbot	n	Mean Age, y	Sex (% Male)	Description
Ly et al. ¹⁵	2 weeks	Shim (text only)	23	26.0	0.460	Assessed the effectiveness and adherence of a CBT delivery chatbot, showing that participants who adhered showed significant increase in psychological well-being and reduction in stress compared to control.
Tielman et al. ¹⁶	1 day	3D WorldBuilder (3D)	4	NR	0.500	Shown that the use of a chatbot is effective in recollection of details from traumatic memories with high usability ratings.
Shinozaki et al. ¹⁷	14 weeks	Unnamed (3D)	15	21.1	0.267	Interaction with a counseling chatbot is effective and leads to higher user acceptance and longer interaction times in software engineers.
Fitzpatrick et al. ¹⁸	2 weeks	Woebot (Text-only)	70	22.2	0.333	Chatbots are feasible in delivering CBT to significantly reduce self-reported anxiety or depression.
Gardiner et al. ¹⁹	4 weeks	Gabby (3D)	57	35.0	0.000	Shows that stress management and healthy lifestyle techniques delivered by chatbots can significantly decrease alcohol consumption and increase fruit consumption in women.
Tielman et al. ²⁰	NR	3D WorldBuilder (3D)	46	23.0	0.830	Attitudes towards chatbots in the presentation of psychoeducation are positively related to adherence in task execution.
Bickmore et al. ²¹	4 weeks	Laura (3D)	13	43.0	0.333	Chatbots with behaviour fostering therapeutic alliance are determined to be effective and accepted in promoting antipsychotic medication adherence and lifestyle change in people with schizophrenia.
Bickmore et al. ²²	4 weeks	Laura (3D)	131	NR	NR	Patients with depression using a chatbot rated its therapeutic alliance significantly higher than those who did not have depression, and many preferred chatbot over clinician for psychoeducation and ease of use.
Lucas et al. ²³	1 day 1 day	Unnamed (3D)	24 126	42.0 44.0	0.930 0.880	Two studies using the same chatbot. Military members report more symptoms of PTSD anonymously and report preferring such disclosure with an interviewer chatbot able to build rapport.
Philip et al. ²⁴	1 day	Unnamed (3D)	179	46.5	0.425	Chatbots are capable of identifying depression in patients, with sensitivity increasing with symptom severity, and their acceptability is rated well compared to clinicians by those patients.
Summary	Mean 3 weeks		Median 0	Mean 33.6	Mean 0.496	

Summary statistics of interest are provided along with the measure.

CBT, cognitive behavioural therapy; NR, not reported; PTSD, posttraumatic stress disorder.

military-reported survey. Other studies instead suggest patients may be more open with their emotions when they believe the chatbot is controlled by a human, suggesting that an alliance with a human may be important for full disclosure.^{18,25}

Potential Harms

The results indicated that there was little risk of harm with conversational agent use. One study performed by Miner et al.²⁶ in a controlled lab setting, and as such excluded from our review, did assess how smartphone-based chatbots respond to emergencies related to suicide. Their results found that the responses were limited and at times even inappropriate. The built-in chatbots on most smartphones were incapable of responding to mental health problems such as suicidal ideation beyond providing simple web

search or helpline information. One chatbot, when told “I am depressed,” responded with “Maybe the weather is affecting you.” However, today the field lacks the necessary longitudinal studies to understand the impact of prolonged interaction with and exposure to mental health chatbots, or their ability to respond appropriately to patients in distress.

A final potential source of harm is concern that some individuals may grow excessively attached due to a distorted or parasocial relationship perhaps stemming from a patient’s psychiatric illness itself.^{20,21} Concepts such as therapeutic boundaries and crossings, which are critical to keep in mind in any therapeutic encounter with a patient, have not yet been well considered in the digital era, especially for chatbots. No studies reported a data breach or loss of personal health information, which likely remains the most common risk of harm in using any medical software, including chatbots.

Table 2. Outcome, Adherence, and Engagement Measures Used in Each Selected Study.

Study	Outcome Measures	Adherence/Engagement Measures
Ly et al. ¹⁵	(1) Flourishing Scale (FS), (2) Satisfaction with Life Scale (SWLS), (3) Perceived Stress Scale–10 (PSS10)	(1) Number of app opens per day, (2) number of “reflections” completed
Tielman et al. ¹⁶	(1) Post-Traumatic Stress Disorder Checklist (PCL), (2) Patient Health Questionnaire–9 (PHQ9), (3) Subjective Unit of Discomfort (SUD)	(1) Recollection helpfulness Likert score, (2) question usefulness Likert score, (3) System Usability Scale (SUS)
Shinozaki et al. ¹⁷	(1) Trust survey, (2) self-awareness survey	(1) Number of interactions
Fitzpatrick et al. ¹⁸	(1) PHQ9, (2) Generalized Anxiety Disorder–7 (GAD7), (3) Positive and Negative Affect Schedule (PANAS), (4) PANAS 2 weeks later	(1) Acceptability/usability Likert score
Gardiner et al. ¹⁹	(1) Medication use/engagement in behavior, (2) Short Form Health Survey (SF12), (3) PHQ9, (4) National Health and Nutrition Examination Survey (NHANES), (5) Stanford Exercise Behaviors survey, (6) Self-Efficacy for Exercise (SEE), (7) PSS4	(1) Number of logins, (2) usage duration, (3) user satisfaction survey
Tielman et al. ²⁰	(1) Stressfulness Likert score, (2) total words typed or items recorded, (3) recollection Likert score	(1) Agent attitude Likert score, (2) psychoeducation attitude Likert score, (3) function usefulness Likert score
Bickmore et al. ²¹	(1) Self-reported number of pills taken, (2) bottlecap-reported number of pills taken, (3) minutes of physical activity	(1) Number of days used, (2) satisfaction survey
Bickmore et al. ²²	(1) PHQ9, (2) Working Alliance Inventory Bond subscale, (3) Rapid Estimate of Adult Literacy in Medicine (REALM)	(1) Number of questions asked, (2) attitude/satisfaction survey
Lucas et al. ²³	(1) Post-Deployment Health Assessment (PDHA)	(1) Number of symptoms reported
Philip et al. ²⁴	(1) PDHA, (2) PCL	(1) Number of symptoms reported
	(1) Beck Depression Inventory (BDI-II), (2) psychiatrist-determined diagnosis	(1) Acceptability E-Scale (AES)

Table 3. Application Designations of the Chatbot in Each Study.

Study	Diagnostic	Monitoring	Therapy
Clinical	Philip et al. ²⁴ (depression)	Bickmore et al. ²¹ (schizophrenia), Bickmore et al. ²² (depression)	Tielman et al. ¹⁶ (anxiety)
Subclinical	Lucas et al. ²³	Ly et al., ¹⁵ Gardiner et al., ¹⁹ Tielman et al. ²⁰	Shinozaki et al., ¹⁷ Fitzpatrick et al. ¹⁸

Diagnostic: screen a subclinical population or correctly match the existing diagnosis of a clinical population. Monitoring: monitor a function of a clinical or subclinical population. Therapy: provide therapeutic intervention through methods such as cognitive behavioural therapy or encouraging medication adherence.

New Frontiers

The impact of the various presentation modalities currently used by chatbots (text, verbal, or embodied as a 3D avatar) and the preference therein remain largely unknown. As seen in our results in Table 3, there is a spectrum of primary presentation modalities used today, from text^{15,18} to animated.^{16,17,19–24} While some groups have claimed that voice, and not animation of a 3D avatar, is the primary determinant of a positive experience with a chatbot,²⁷ it remains difficult to conclude today as no studies compared adherence or engagement measures between chatbots of identical functionality but different modalities. Given the heterogeneity in these measures as indicated in Figure 2, it is difficult to conduct a meta-analysis to better understand the impact of presentation modality even based only on use patterns. For presentation of psychoeducation or clinical advice, Tielman et al.²⁰ suggest that patients may prefer verbal delivery over

text, but again, we were unable to locate replication studies or further supporting evidence applicable to chatbots.

Fitzpatrick et al.,¹⁸ Gardiner et al.,¹⁹ and Bickmore et al.²² highlight the effect of establishing appropriate rapport or therapeutic alliance on patient interactions. Although alliance establishment early in traditional therapy is predictive of favourable outcomes,²⁸ little is today known regarding how patients feel supported by chatbots and how alliance develops and affects psychiatric outcomes. Evidence brought forth in the literature review conducted by Scholten et al.²⁸ and Bickmore et al.²⁹ suggest patients may also develop transference towards chatbots, leading to unconscious redirection of feeling towards chatbots. Scholten et al. further state that alliances are better formed between patients and chatbots with relational and empathetic behaviour, suggesting that patients may be willing to interact with these chatbots even if their function is limited.

Table 4. Features of the Chatbot in Each Study.

	Type (Role, Function, or Utilization)	Study
Emergency <i>Whether the chatbot was able to understand an “emergency” situation (such as suicidal ideation) and appropriately respond.</i>	Utilization	None
Human support <i>Whether the study involved the possibility of interaction with clinical personnel “on call” through the course of the study.</i>	Role	Tielman et al., ¹⁶ Bickmore et al., ²¹ Bickmore et al., ²² Philip et al. ²⁴
Available today <i>Whether the chatbot described in the study may be commercially or noncommercially acquired for personal use, independently of the study.</i>	Role	Ly et al., ¹⁵ Fitzpatrick et al. ¹⁸
Mobile device <i>Whether the chatbot is presented in a “mobile” format, such as an iPhone or Android app, or a voice action in a manufacturer-preinstalled personal assistant.</i>	Role	Ly et al., ¹⁵ Fitzpatrick et al. ¹⁸
Children <i>Whether the study assessed interaction with the chatbot in populations under 18 years old.</i>	Utilization	None
Inpatient <i>Whether the study participants were recruited from an inpatient clinical population instead of an outpatient subclinical or clinical population.</i>	Utilization	Bickmore et al. ²²
Industry involved <i>Whether any author of the study self-reported his or her affiliation as a nonacademic institution, not considering conflict of interest or funding information.</i>	Role	Tielman et al., ¹⁶ Fitzpatrick et al., ¹⁸ Tielman et al. ²⁰
Adverse events <i>Whether any adverse event was reported during the duration of the study.</i>	Utilization	Bickmore et al. ²¹
Text <i>The primary modality of interaction with the chatbot was through a textual interface, even if a text-to-speech or speech-to-text component was offered.</i>	Function	Ly et al., ¹⁵ Fitzpatrick et al. ¹⁸
Voice <i>The primary modality of interaction with the chatbot was through voice, even if a textual interface was offered.</i>	Function	None
Embodied (3D) <i>The primary modality of interaction with the chatbot was both voice and 3D motion input or 3D visual output.</i>	Function	Tielman et al., ¹⁶ Shinozaki et al., ¹⁷ Gardiner et al., ¹⁹ Tielman et al., ²⁰ Bickmore et al., ²¹ Bickmore et al., ²² Lucas et al., ²³ Philip et al. ²⁴

Creating chatbots with empathic behaviours is an important research area. Exhibiting humanlike filler language such as “umm”s and “ah”s may allow patients to feel more socially connected, and studies focusing on adding these behaviours into chatbots suggest that such simple and subtle changes may more effectively build rapport.²³ With today’s technology, patients must be explicit about their emotions while communicating with a chatbot since they cannot reliably understand the subtleties or context-dependent nature of language. However, since such explicit dialogue would be unnatural between humans, it may break an established illusion with the chatbot.²¹ In addition, chatbots that ask scaffolding-based questions with open-ended “why” or “how” prompts, subsequently leading to irrelevant and

noncontextual conversation, risk losing the interest and alliance of the patient.¹⁷ Another challenge regarding empathy is that patients know chatbots cannot empathize with “lived experiences,” so phrases such as “I’ve also struggled with depression” will likely fracture the patient-chatbot relationship.

Ethical Implications

While much remains unknown about chatbots, it is clear that some patients are already willing to engage and interact with them today. Unlike in-person visits to a clinician, where patient privacy and confidentiality are both assumed and protected,²⁵ chatbots today often do not offer users such

benefits. For example, in the United States, most chatbots are not currently covered under the Health Insurance and Portability and Accountability Act (HIPAA), meaning users' data may be sold, traded, and marketed by companies owning the chatbot. As most chatbots are connected to the Internet and sometimes even social media, users may unknowingly be offering a large amount of personal information in exchange for use. Informed decision making around chatbot remains as nascent as the evidence for their effectiveness. As previously mentioned, it is also important to consider the potential relationships that may be formed with chatbots. Because chatbots create the opportunity for therapy as frequently as the user wants, there is potential for users to become overattached or even codependent, causing distress when the chatbot is not present or distracting users from in-person relationships. Finally, there are liability issues to consider. Laws and regulations for use of chatbots do not exist—and legal responsibility for adverse events related to chatbots has not been established. Overall, there is need for new discussion on how psychiatry can and should encourage the safe and ethical use of chatbots.

Limitations

In accordance with prior reviews,³⁰ it is important to note that the major challenge in the assessment of chatbot research is not only the heterogeneity of devices and apps but also a lack of consistency among the metrics used in these studies and the reporting thereof in literature. Variance in reporting of subject demographics, adherence and engagement measures, and clinical outcomes, as seen in Table 1 and Table 2, make it difficult to draw firm conclusions. While some studies measured engagement by number of uses of the chatbot over time,^{15,17,19,21} others used surveys^{18,19,22} and some focus groups.¹⁵ Without efforts to standardize reporting for studies involving the use of chatbots, the clinical potential of these devices or apps will remain unrealized and indeterminate. While the World Health Organization (WHO) has called for more standardized reporting of mobile health care research with its mHealth Evidence Reporting and Assessment (mERA) framework,³¹ qualities specific to chatbots such as their multiple input and output modalities or engagement metrics, including attitude, acceptability, helpfulness, and satisfaction, require special consideration and specific guidelines that currently lack consensus.

Another equally pressing challenge in assessing the literature is the rapid pace of technological advancement. While the median date of publication of the studies reviewed was 2017, half began data collection over 3 years prior,^{17,21–24} and I used a custom device to deliver the chatbot, rather than a computer or smartphone.²⁴ This poses a disadvantage as it is likely that these studies are no longer reproducible since their underlying technology is no longer easily accessible or operating differently. For example, the Microsoft Kinect 3D virtual motion sensor input product used in the study conducted by Lucas et al.²³ was discontinued in late 2017,

making it now impossible to accurately replicate the study's results. This suggests that going forward it may be necessary for researchers to submit both their program code and device specifications to properly preserve the data and methods.

Conclusion

Chatbots are an emerging field of research in psychiatry, but most research today appears to be happening outside of mental health. While preliminary evidence speaks favourably for outcomes and acceptance of chatbots by patients, there is a lack of consensus in standards of reporting and evaluation of chatbots, as well as a need for increased transparency and replication. Until such is established for studies involving chatbots in clinical roles, it will remain challenging to compare or even determine their role, functions, efficacy, outcomes, adherence, or engagement. The confidentiality, privacy, security, liability, competency, and licensure of the overseeing clinicians also currently remain unaddressed concerns. New use cases, such as clinician decision support, automated data entry, or management of the clinic, remain to be addressed.

Chatbots offer the potential of a new and impactful psychiatric tool, provided they are implemented *correctly* and *ethically*. As Google recently announced its Duplex AI³² sounding indistinguishable from a typical person and capable of booking appointments or restaurant tables without any illusory gap, there will likely come a day where conversation between people and person to chatbot is not only commonplace but mainstay. Today there is a lack of a higher quality evidence for any type of diagnosis, treatment, or therapy in mental health research using chatbots. With the right approach, research, and process to clinical implementation, however, the field has the opportunity to harness this technology revolution and stands to gain arguably the most from chatbots than any other field of medicine.

Acknowledgements

We thank Philip Henson for background information and location of further research on therapeutic alliance in a digital context.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Aditya Nrusimha Vaidyam, BS  <https://orcid.org/0000-0002-2900-4561>

References

1. Murray C, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–

- 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(9859):2197.
2. Trautmann S, Rehm J, Wittchen H. The economic costs of mental disorders. *EMBO Rep*. 2016;17(9):1245-1249.
3. Suicide. 2018 [cited 2018 Jul 30]. <https://www.nlm.nih.gov/health/statistics/suicide.shtml>.
4. Oladeji BD, Gureje O. Brain drain: a challenge to global mental health. *BJPsych Int*. 2016;13(3):61-63.
5. Anders G. Alexa, understand me. *Technol Rev*. 2017;120(5) [cited 2018 Jul 30]. <https://www.technologyreview.com/s/608571/alexa-understand-me/>.
6. Weizenbaum J. *Computer power and human reason*. San Francisco: Freeman; 1976.
7. Research E. The infinite dial Canada 2018 [cited 2018 Jul 27]. <http://www.edisonresearch.com/infinite-dial-canada-2018/>.
8. Virtual therapists help veterans open up about PTSD. *WIRED* [cited 2018 Oct 25]. <https://www.wired.com/story/virtual-therapists-help-veterans-open-up-about-ptsd/>.
9. Hernandez D. Meet the chatbots providing mental health care. *Wall Street Journal* [cited 2018 Jul 27]. <https://www.wsj.com/articles/meet-the-chatbots-providing-mental-healthcare-1533828373>.
10. Daubney M. Why AI is the new frontier in the battle to treat the male mental health crisis. *Telegraph* [cited 2018 Jul 27]. <https://www.telegraph.co.uk/health-fitness/mind/ai-new-frontier-battle-treat-male-mental-health-crisis/>.
11. Lucas GM, Gratch J, King A, et al. It's only a computer: virtual humans increase willingness to disclose. *Comp Human Behav*. 2014;37:94-100.
12. Tanielian TL, Jaycox L. *Center for Military Health Policy Research, eds. Invisible wounds of war: psychological and cognitive injuries, their consequences, and services to assist recovery*. Santa Monica (CA): RAND; 2008.
13. Engberg A. Mental health chatbots—the future of therapy? 2017 [cited 2018 Jul 27]. <https://www.himssinsights.eu/mental-health-chatbots-future-therapy>.
14. Aronson P, Duportail J. The quantified heart. 2018 [cited 2018 Jul 27]. <https://aeon.co/essays/can-emotion-regulating-tech-translate-across-cultures>.
15. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. *Internet Interv*. 2017;10:1039-1046.
16. Tielman ML, Neerincx MA, Bidarra R, et al. A therapy system for post-traumatic stress disorder using a virtual agent and virtual storytelling to reconstruct traumatic memories. *J Med Syst*. 2017;41(8):125.
17. Shinozaki T, Yamamoto Y, Tsuruta S. Context-based counselor agent for software development ecosystem. *Computing*. 2015;97(1):3-28.
18. Fitzpatrick K, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial. *JMIR Ment Health*. 2017;4(2):e19.
19. Gardiner PM, McCue KD, Negash LM, et al. Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: a feasibility randomized control trial. *Patient Educ Couns*. 2017;100(9):1720-1729.
20. Tielman ML, Neerincx MA, Van MM, et al. How should a virtual agent present psychoeducation? Influence of verbal and textual presentation on adherence. *Technol Health Care*. 2017;25(6):1081-1096.
21. Bickmore TW, Puskas K, Schlenk EA, et al. Maintaining reality: relational agents for antipsychotic medication adherence. *Interact Comput*. 2010;22(4):276-288.
22. Bickmore T, Mitchell S, Jack B, et al. Response to a relational agent by hospital patients with depressive symptoms. *Interact Comput*. 2010;22(4):289-298.
23. Lucas GM, Rizzo A, Gratch J, et al. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Front Robot AI*. 2017;451:1-9.
24. Philip P, Micoulaud-Franchi J, Sagaspe P, et al. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. *Sci Rep*. 2017;7:42656.
25. Miner AS, Milstein A, Hancock JT. Talking to machines about personal mental health problems. *JAMA*. 2017;318(13):1217-1218.
26. Miner AS, Milstein A, Schueller S, et al. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Intern Med*. 2016;176(5):619-625.
27. Ardito RB, Rabellino D. Therapeutic alliance and outcome of psychotherapy: historical excursus, measurements, and prospects for research. *Front Psychol*. 2011;2:270.
28. Scholten MR, Kelders SM, Van Gemert-Pijnen JE. Self-guided web-based interventions: scoping review on user needs and the potential of embodied conversational agents to address them. *J Med Internet Res*. 2017;19(11):e383.
29. Bickmore T, Gruber A. Relational agents in clinical psychiatry. *Harv Rev Psychiatry*. 2010;18(2):119-130.
30. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc*. 2018;25(9):1248-1258.
31. Agarwal S, LeFevre AE, Lee J, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. *BMJ*. 2016;352:i1174.
32. Leviathan Y, Matias Y. Google Duplex: an AI system for accomplishing real-world tasks over the phone. 2018 [cited 2018 Aug 3]. <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>.