

Changes to the Psychiatric Chatbot Landscape: A Systematic Review of Conversational Agents in Serious Mental Illness

The Canadian Journal of Psychiatry /
La Revue Canadienne de Psychiatrie
2021, Vol. 66(4) 339-348
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0706743720966429
TheCJP.ca | LaRCP.ca



Changements du paysage psychiatrique des chatbots: une revue systématique des agents conversationnels dans la maladie mentale sérieuse

Aditya Nrusimha Vaidyam, MS¹ , Danny Linggonegoro, BS¹, and John Torous, MD, MBI¹ 

Abstract

Objective: The need for digital tools in mental health is clear, with insufficient access to mental health services. Conversational agents, also known as chatbots or voice assistants, are digital tools capable of holding natural language conversations. Since our last review in 2018, many new conversational agents and research have emerged, and we aimed to reassess the conversational agent landscape in this updated systematic review.

Methods: A systematic literature search was conducted in January 2020 using the PubMed, Embase, PsychINFO, and Cochrane databases. Studies included were those that involved a conversational agent assessing serious mental illness: major depressive disorder, schizophrenia spectrum disorders, bipolar disorder, or anxiety disorder.

Results: Of the 247 references identified from selected databases, 7 studies met inclusion criteria. Overall, there were generally positive experiences with conversational agents in regard to diagnostic quality, therapeutic efficacy, or acceptability. There continues to be, however, a lack of standard measures that allow ease of comparison of studies in this space. There were several populations that lacked representation such as the pediatric population and those with schizophrenia or bipolar disorder. While comparing 2018 to 2020 research offers useful insight into changes and growth, the high degree of heterogeneity between all studies in this space makes direct comparison challenging.

Conclusions: This review revealed few but generally positive outcomes regarding conversational agents' diagnostic quality, therapeutic efficacy, and acceptability, which may augment mental health care. Despite this increase in research activity, there continues to be a lack of standard measures for evaluating conversational agents as well as several neglected populations. We recommend that the standardization of conversational agent studies should include patient adherence and engagement, therapeutic efficacy, and clinician perspectives.

Abrégé

Objectif : Le besoin d'instruments numériques en santé mentale est évident, car l'accès aux services de santé mentale est insuffisant. Les agents conversationnels, aussi appelés chatbots ou assistants vocaux, sont des instruments numériques capables de tenir des conversations en langage naturel. Depuis notre dernière revue en 2018, bien des agents conversationnels et de nombreuses études ont émergé, et nous avons cherché à réévaluer le paysage des agents conversationnels dans cette revue systématique mise à jour.

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

Corresponding Author:

John Torous, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, Massachusetts 02215, USA.
Email: jtorous@bidmc.harvard.edu

Méthodes : Une recherche systématique de la littérature a été menée en janvier 2020, dans les bases de données PubMed, EmBase, PsychINFO, et Cochrane. Les études comprenaient celles qui portaient sur un agent conversationnel évaluant une maladie mentale sérieuse : trouble dépressif majeur, troubles du spectre de la schizophrénie, trouble bipolaire ou trouble anxieux.

Résultats : Sur les 247 références identifiées dans les bases de données choisies, sept études satisfaisaient aux critères d'inclusion. Globalement, les agents conversationnels affichaient des expériences généralement positives en ce qui concerne la qualité diagnostique, l'efficacité thérapeutique, ou l'acceptabilité. Il y a encore cependant une absence de mesures normalisées qui permettent de facilement comparer les études dans ce contexte. Plusieurs populations manquaient de représentation, notamment la population pédiatrique et celle souffrant de la schizophrénie ou du trouble bipolaire. Même si comparer la recherche de 2018 avec celle de 2020 offre un aperçu utile sur les changements et la croissance, le degré élevé d'hétérogénéité entre toutes les études dans ce domaine rend la comparaison directe difficile.

Conclusions : Cette revue a révélé des résultats peu nombreux mais généralement positifs à l'égard de la qualité diagnostique, de l'efficacité thérapeutique, et de l'acceptabilité des agents conversationnels, qui peuvent s'ajouter aux soins de santé mentale. Malgré l'accroissement de l'activité de recherche, il manque encore de mesures normalisées pour évaluer les agents conversationnels ainsi que plusieurs populations négligées. Nous recommandons que la normalisation des études sur les agents conversationnels inclue l'observance et l'engagement des patients, l'efficacité thérapeutique, et les perspectives du clinicien.

Keywords

conversational agent, chatbot, mental health, serious mental illness, psychiatry

Introduction

The need for digital tools in mental health is clear, with insufficient access to mental health services worldwide and clinical staff increasingly unable to meet rising demand.¹ The World Health Organization (WHO) reports that depression is a leading cause of disability worldwide.² In 2013, the economic cost of treatment for mental health and substance abuse disorders in the United States alone was nearly US\$200 billion.³ Contributing to mental health vulnerability and burden, social isolation and loneliness has been quantified as an epidemic associated with an estimated US\$6.7 billion in treatment.⁴ Response to emergent global situations such as COVID-19 is exacerbating the already-limited access to services, and the urgent need for innovation around access to mental health care has become clear. Innovation offered via artificial intelligence and conversational agents has been proposed as one means to increase access and quality of care.^{5,6}

Conversational agents, also known as chatbots or voice assistants, are digital tools capable of holding natural language conversations and mimicking human-like behavior in task-oriented dialogue with people. Conversational agents exist in the form of hardware devices such as Amazon Echo or Google Home as well as software apps such as Amazon Alexa, Apple Siri, and Google Assistant. It is estimated that 42% of U.S. adults use digital voice assistants on their smartphone devices,⁷ and some industry studies claim that nearly 24% of U.S. adults own at least 1 smart speaker device.⁸ With such widespread access to conversational agents, it is understandable that many are interested in their potential and role in health care. Early research has explored the use of conversational agents in a diverse range of clinical settings

such as helping with diagnostic decision support,⁹ education regarding mental health,¹⁰ and monitoring of chronic conditions.¹⁰ In 2018, the most common condition that conversational agents claimed to cover was related to mental health.¹⁰ Other conditions included hypertension, asthma, type 2 diabetes, obstructive sleep apnea, sexual health, and breast cancer.¹⁰

In today's evolving landscape of rapidly changing technology, growing global health concerns, and lack of access to high-quality mental health care, use and evaluation of these conversational agents of mental health continue to evolve. Since our team's 2018 review on the topic, many new conversational agents, products, and research studies have emerged.¹¹ While conversational agents have been posited to benefit patients and providers, many risks of conversational agents such as possibly disrupting the therapeutic alliance have not been fully elucidated.¹² As emerging research in suicide prevention is evaluating the use of automated processes to detect risk—the need for a clear understanding of the current state of the field is patent. Some conversational agents offer self-help programs to reduce anxiety and depression,¹³ while others are continually being assessed for use as diagnostic aids¹⁴ with goals of entering into clinical settings.

Like any digital tool, conversational agents raise concerns and complications surrounding privacy breaches and lack of guidance on regulatory/legal duties.¹¹ Individuals may be using conversational agents and other digital tools with or without recommendation from their physician or psychiatrist. Clinicians need to be aware of what the current evidence and actual abilities of these conversational agents are rather than relying on company marketing materials, which may offer a biased and overvalued estimation. To understand

which conversational agents are effective and their potential uses—as well as harms and risks—researchers must continue to correctly understand the kinds of effects they may posit within and outside of the clinical mental health setting.

Despite this increase in the availability of conversational agents, in our prior review, we found a lack of higher quality evidence for any type of diagnosis, treatment, or therapy in published mental health research using conversational agents.¹⁵ Another review in 2019 reported only a single randomized controlled trial measuring the efficacy of conversational agents in general.¹⁰ It is also important to consider the role of industry-sponsored studies and potential adverse impacts this may have on study design.¹⁶ Our prior preliminary systematic review of the landscape was performed in 2018,⁸ finding high heterogeneity regarding outcomes and reporting metrics for the conversational agents, missingness of information, crucial factors such as engagement, adverse events, and more. Now 2 years later and given the advent of new research, products, and claims, in this review, we revisit the flourishing mental health conversational agent space, quantifying these factors and suggesting novel or alternative approaches where appropriate.

Methods

For this comparative review, the same terms as defined in our prior systematic review on conversational agents were used; that is, a combination of keywords including “conversational agent” or “chatbot” without other filter parameters for peer-reviewed published papers (and not conference proceedings, poster abstracts, etc.) in the English language only between July 2018 and January 2020. These terms were selected initially as they provided the largest set of relevant articles. This literature search was conducted in January 2020 on the same select databases (PubMed, Embase, PsycINFO, and Cochrane), with the exception of Web of Science and IEEE Xplore, primarily due to a lack of clinically relevant literature, as determined in our prior review. Title, abstract, and full-text screening and full-text data extraction phases were conducted by 2 authors (A.N.V and D.W.L) through discussion. Disagreements in screening phases were resolved through discussion and majority consensus. Reasons for exclusion were compiled. From included articles, data extraction comprised of study characteristics (duration of study, conversational agent name, ability for unconstrained natural language, sample size, mean age, sex), study outcomes and engagement measures, and conversational agent features. Ability for unconstrained natural language was also assessed, as this feature could pose serious safety risks.¹⁷

Studies were selected that measured the effect of conversational agents on patients with serious mental illness (SMI): major depressive disorder, schizophrenia spectrum disorders, bipolar disorder, or anxiety disorder, either diagnosed or self-reported. SMI was chosen as the primary population of focus due to noticeable trends in these areas, whereas conversational agents in other psychiatric disorders with

lower prevalence have not been substantially studied. Notably, substance abuse disorders were included in the prior review and are here excluded in order to narrow our focus, as the use of conversational agents in this population may merit its own in-depth review today. Given that conversational agents have many forms, it was ensured that any selected study was agreed upon by all authors.

Studies were excluded if the study protocol did not measure the direct effect of the use of a conversational agent or did not at all involve the conversational agent in diagnosis, management, or treatment of SMI. Studies were also excluded if the conversational agent used by participants according to the study protocol did not dynamically generate its content through natural language processing; for example, “Wizard of Oz”-style conversational agents that match input dialogue or query to recycled statements from other users did not qualify. Abstracts, reviews, and ongoing clinical trials were excluded. Non-English-language manuscripts were excluded.

Results

Of the 247 new references identified from search terms applied to selected databases, 65 duplicates were removed and 154 were screened out based on title and abstract. Of the 28 studies identified for full-text screening, 11 were not relevant to our aims related to study population. Only 7 studies were identified as relevant for the data extraction phase. Figure 1 depicts the detailed Preferred Reporting Item for Systematic Reviews and Meta-Analyses (PRISMA) diagram.

Summary data for the 7 studies in Table 1 are relatively similar to our prior review in 2018. The mean age of participants in the study was 34.29. The mean number of participants was 74. The mean study duration was 4.6 weeks. Measures examined are included in Table 2. Notably, only 4 of 7 studies included adherence or engagement measures. In comparison, all studies in 2018 had some level of engagement measure. Table 3 shows clinical targets of each study; 4 studies investigated major depressive disorder, and 3 examined anxiety disorders. While our 2018 review contained a single study that investigated schizophrenia,¹⁸ schizophrenia and bipolar disorders were not examined in any of the studies in this review.

Overall, features of the conversational agents in Table 4 were mostly unchanged in comparison to 2018. Similar to our prior review, there continued to be no inclusion of children or consideration for emergency situations in these studies as well as minimal reporting of adverse effects. Notably, more conversational agents examined within the past 2 years were available on text and mobile device interface rather than other modalities (such as 3D avatars, with or without motion output) as was found in our 2018 review.

Discussion

Conversational agents continue to gain interest, given their potential to expand access to mental health care. In this

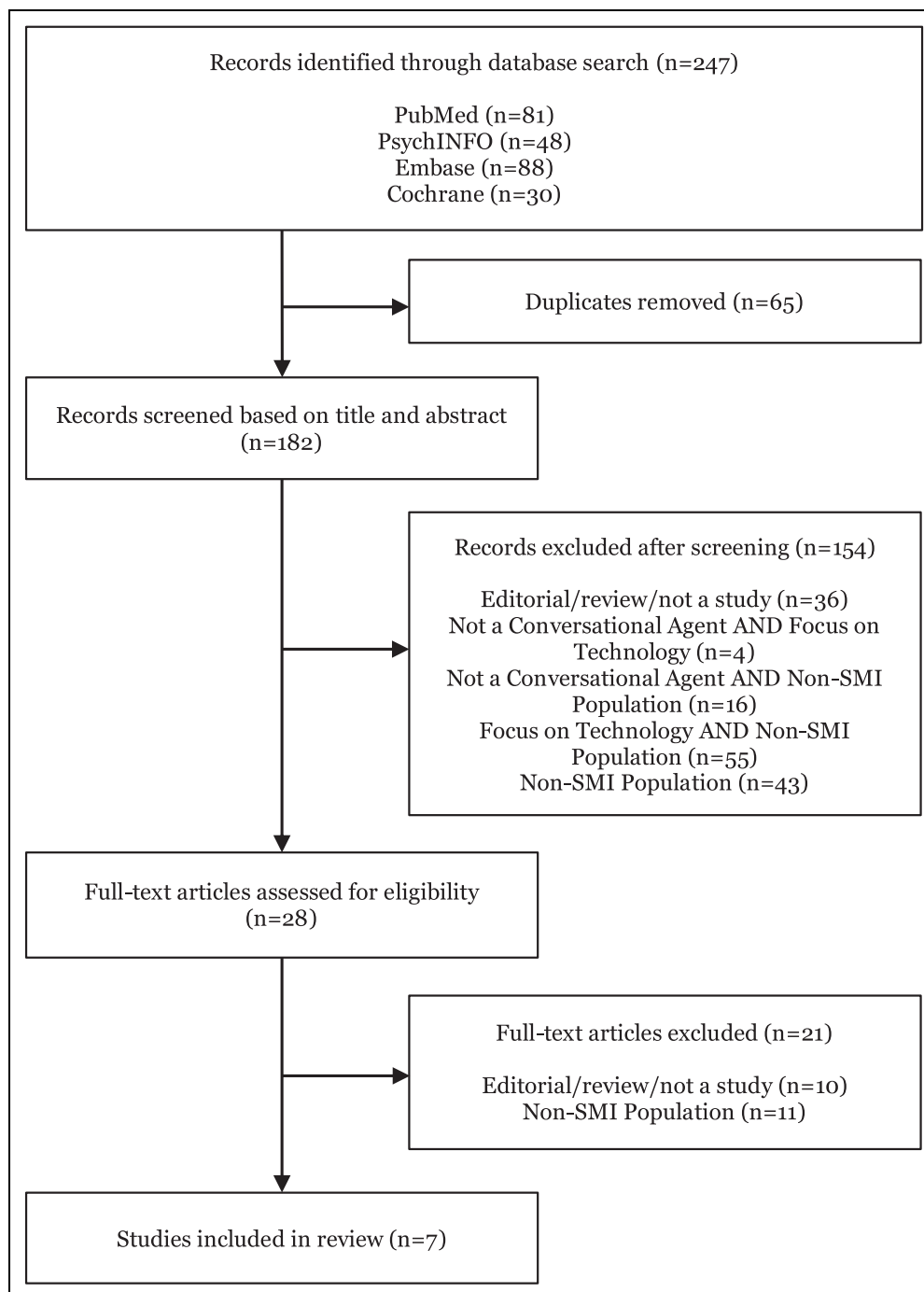


Figure 1. Preferred Reporting Item for Systematic Reviews and Meta-Analyses (PRISMA) diagram.

updated review, 7 new studies were included. Two of these studies focused on assessing diagnostic quality, 3 studies examined therapeutic efficacy, and 2 studies evaluated the acceptability. Compared to our prior review in 2018, among the 7 new studies, there continues to be no consistent measure to evaluate engagement, yet more conversational agents are now available on mobile devices. Over half of the conversational agents were focused on depression, while schizophrenia and bipolar disorders had no representation in research output in the last 2 years.

Diagnostic Quality, Therapeutic Efficacy, and Acceptability

Two new studies focused on the diagnostic quality of the conversational agents when compared to a gold standard. Jungmann et al. compared the diagnosis of mental disorders in the conversational agent Ada versus psychotherapists, psychology students, and laypersons and concluded that the conversational agent had high diagnostic agreement with psychotherapists and moderate diagnostic agreement with

Table 1. Reported Information about Each Selected Study and Conversational Agent Is Provided.

Authors (Year)	Duration	Conversational Agent	Unconstrained Input	n	Mean Age	Sex (%M)
Fulmer et al. ¹⁹ (2018)	2 and 4 Weeks	Tess	Yes	74	22.9	.28
Description: Assessed the feasibility and efficacy of reducing self-identified symptoms of depression and anxiety in college students. Randomized controlled trial showed reduction in symptoms of depression and anxiety.						
Inkster et al. ²⁰ (2018)	2 Months	Wysa	Yes	129	N/A	N/A
Description: Evaluated effectiveness and engagement levels of AI-enabled empathetic, text-based conversational mobile mental well-being app on users with self-reported symptoms of depression; 67% of users found the app helpful and encouraging. More engaged users had significantly higher average mood improvement compared to low users.						
Jungmann et al. ²¹ (2019)	3 to 6 Hours	Ada	No	6	34	.5
Description: Investigated diagnostic quality of a health app for a broad spectrum of mental disorders and its dependence on expert knowledge. Psychotherapists had a higher diagnostic agreement (in adult cases) between the main diagnosis of a case vignette in a textbook and the result given by the app.						
Martínez-Miranda et al. ²² (2019)	8 Weeks	HelPath	No	18	31.5	.63
Description: Assessed acceptability, perception, and adherence toward the conversational agent. Participants perceived embodied conversational agents as emotionally competent, and a positive level of adherence was reported.						
Philip et al. ²³ (2020)	N/A	Unnamed virtual medical agent	No	318	45.01	.45
Description: Measured engagement and perceived acceptance and trust of the virtual medical agent in diagnosis of addiction and depression. Although 68.2% of participants reported being very satisfied with the virtual medical agent, only 57.23% were willing to interact with the virtual medical agent in the future.						
Provoost et al. ²⁴ (2019)	N/A	Sentiment mining algorithm tailored toward Dutch language	Yes	52	N/A	N/A
Description: Evaluated accuracy of automated sentiment analysis against human judgment. Online cognitive behavioral therapy patient's user texts were evaluated on an overall sentiment and the presence of 5 specific emotions by an algorithm and by psychology students. Results showed moderate agreement between the algorithm and human judgment when evaluating overall sentiment (positive vs. negative sentiment) but low agreement with specific emotions.						
Suganuma et al. ²⁵ (2018)	1 Month, minimum 15 days	SABORI	No	454	38.04	.308
Description: Evaluated feasibility and acceptability study of a conversational agent. Results showed improvement in WHO-5 and Kessler 10 scores.						
Mean: 4.6 weeks			Median: 74 Mean: 34.29 Mean: 43.36%			

Note. Unconstrained input refers to whether the user of the conversational agent is able to freely converse with the agent, as opposed to constrained input where specific dialogue options are presented from which the user must select a choice to continue the conversation. N/A = not available; AI = artificial intelligence; WHO = World Health Organization.

psychology students and laypersons.²¹ The conversational agent had lower diagnostic agreement with all participants in child and adolescent cases, which suggests that pediatric cases may be more nuanced. Provoost et al. compared the accuracy of an automated sentiment analysis against human judgment. User texts were evaluated on overall sentiment and the presence of specific emotions detected by an algorithm and psychology students. Results showed moderate agreement between algorithm and human judgment in evaluating overall sentiment (either positive or negative); however, there was low agreement with specific emotions such as pensiveness, annoyance, acceptance, optimism, and serenity. These results suggest that there continues to be room for improvement in the diagnostic quality of these particular conversational agents.

Three studies examined the therapeutic efficacy of different conversational agents. Fulmer et al. found that the conversational agent Tess was able to reduce self-identified symptoms of depression and anxiety in college students.¹⁹ Inkster et al. studied the conversational agent Wysa and found users who were more engaged with the conversational agent had significantly higher average mood improvement compared to lower engagement users.²⁵ Suganuma et al. found that the conversational agent SABORI was effective in improving metrics on WHO-5, a measure of well-being, and Kessler 10, a measure of psychological distress on the anxiety–depression spectrum.²⁰ While these results show promise, the effect of conversational agents as an adjunct to in-person psychiatric treatment has been understudied.¹² It is unclear whether these results are

Table 2. Outcome and Engagement Measures Used.

2020 Studies	Outcome Measures	Engagement Measurement
Fulmer et al. ¹⁹	(1) Depression (PHQ-9) (2) Anxiety (GAD-7) (3) Positive and Negative Affect Scale (4) User satisfaction (survey)	(1) Number of messages exchanged between the participant and the conversational agent, Tess, compared to the participant and the e-book
Inkster et al. ²⁰	(1) Self-reported PHQ-9	(1) Engagement effectiveness: User's in-app feedback responses were performed using thematic analysis (2) Engagement efficacy: Analysis of objections raised by users—conversation messages were tagged for “objection” or “no objection.” Objections were either refusals (user says: “I don't want to do this” to a bot's understanding of what was said) or complaints (“That's not what I said” to a bot's response)
Jungmann et al. ²¹	Agreement between main diagnosis of case vignette in textbook and result given by the app	Not specified
Martínez-Miranda et al. ²²	(1) Hamilton Depression Rating, using the validated Spanish version (2) Plutchik Suicide Risk Scale, validated Spanish version (3) Hamilton Anxiety Rating Scale, validated Spanish version	(1) Average number of sessions carried by users with ECA (2) Total duration in minutes
Philip et al. ²³	(1) 12 Survey questions regarding credibility, benevolence, satisfaction, and usability	(1) Surveyed “are you willing to engage in a new interaction with the virtual agent?” after the interview with the virtual medical assistant to assess future engagement
Provoost et al. ²⁴	(1) Agreement between algorithm and human judgment	Not specified
Suganuma et al. ²⁵	(1) WHO-5 score (2) Kessler 10 score (3) BADS	Not specified

Note. WHO = World Health Organization; GAD-7 = General Anxiety Disorder-7; BADS = Behavioral Activation for Depression Scale; PHQ-9 = Patient Health Questionnaire-9; ECA = embodied conversational agent (chatbot).

Table 3. Conversational Agent Clinical Targets in Present and Prior Reviews.

Type of Severe Mental Illness	Since 2018 Review	2018 Review
Major depressive disorder	Fulmer et al., ¹⁹ Inkster et al., ²⁰ Philip et al., ²³ Suganuma et al. ²⁵	Fitzpatrick et al., ²⁶ Bickmore et al., ²⁷ Philip et al. ¹⁴
Anxiety	Fulmer et al., ¹⁹ Martínez-Miranda et al., ²² Suganuma et al. ²⁵	Fitzpatrick et al., ²⁶ Lucas et al. ²⁸ (PTSD), Tielman et al. ²⁹ (PTSD)
Schizophrenia	None	Bickmore et al. ¹⁸
Bipolar	None	None

Note. PTSD = post-traumatic stress disorder.

generalizable and applicable to a broader population, given inadequate participant characterization in the included studies. Further research is required to determine appropriate indications for the use of adjunctive conversational agents.

Two studies specifically evaluated the acceptability of the conversational agents by patients. Martínez et al. assessed acceptability, perception, and adherence of users toward HelPath, a conversational agent that is used to detect suicidal behavior.²² Participants perceived HelPath as emotionally competent and reported a positive level of adherence. Philip et al. found that the majority (68.2%) of patients rated the virtual medical agent positively very satisfied; 68.2% of

patients “totally agreed” that the virtual medical agent was benevolent, and 79.2% rated the virtual medical agent more than 66% for credibility. Interestingly, despite nearly 66% of patients were “very satisfied,” only 57.23% were willing to interact with the virtual medical agent again in the future, which highlights the continuing need to measure engagement and adherence in future studies.

Heterogeneity in Reporting Metrics

There were few improvements in the standardization of conversational agent evaluation in this review compared to our

Table 4. Conversational Agent Extracted Metrics in Present and Prior Reviews.

Features	Since 2018 Review	2018 Review
Emergency Whether the conversational agent was able to understand an emergency situation and appropriately respond	None	None
Human support Whether the study involved the possible interaction with clinical personnel “on call” through the course of the study	None	Tielman et al., ²⁹ Bickmore et al., ¹⁸ Bickmore et al., ²⁷ Philip et al. ¹⁴
Available today Whether the conversational agent described in the study may be commercially or noncommercially acquired for personal use, independently of the study	All	Ly et al., ³⁰ Fitzpatrick et al. ²⁶
Mobile device Whether the conversational agent is presented in a “mobile” format such as an iPhone or Android app or in a manufacturer-preinstalled personal assistant	Fulmer et al., ¹⁹ Inkster et al., ²⁰ Jungmann et al., ²¹ Martínez-Miranda et al., ²² Philip et al., ²³ Suganuma et al. ²⁵	Ly et al., ³⁰ Fitzpatrick et al. ²⁶
Children Whether the study assessed interaction with the conversational agent in populations under 18 years old.	None; however, Jungmann et al. ²¹ examined prewritten pediatric vignettes	None
Inpatient Whether the study participants were recruited from an inpatient clinical population	None	Bickmore et al. ²⁷
Industry involved Whether any author of the study self-reported his or her affiliation as a nonacademic institution	Fulmer et al., ¹⁹ Inkster et al. ²⁰	Tielman et al., ²⁹ Fitzpatrick et al., ²⁶ Tielman et al. ³¹
Adherence/engagement Whether any adherence or engagement measures were reported during the duration of the study.	Philip et al., ²³ Martínez-Miranda et al., ²² Inkster et al., ²⁰ Fulmer et al. ¹⁹	All
Adverse events Whether any adverse event was reported during the duration of the study	None	Bickmore et al. ¹⁸
Text The primary modality of interaction with the conversational agent was through a textual interface	Fulmer et al., ¹⁹ Inkster et al., ²⁰ Jungmann et al., ²¹ Martínez-Miranda et al., ²² Provoost et al., ²⁴ Suganuma et al. ²⁵	Ly et al., ³⁰ Fitzpatrick et al. ²⁶
Voice The primary modality of interaction with the conversational agent was through voice, even if a textual interface was offered	Philip et al. ²³	None
3D animated The conversational agent incorporated 3D motion output	Martínez-Miranda et al., ²² Philip et al. ²³	Tielman et al., ²⁹ Shinozaki et al., ³² Gardiner et al., ³³ Tielman et al., ³¹ Bickmore et al., ¹⁸ Bickmore et al., ²⁷ Lucas et al., ²⁸ Philip et al. ¹⁴

prior 2018 review. From the 7 included studies, a continued heterogeneity of conversational agents and reported metrics is present, and none of the studies measured engagement in the same way. Fulmer et al. measured engagement as the number of messages exchanged between the participant and the conversational agent,¹⁹ Martínez-Miranda et al. used duration of time spent,²² and Philip et al. surveyed participants regarding engagement.²³ Given the growing literature surrounding the use of conversational agents in health care,

the development of standardized methods of collecting and reporting data is imperative, without which the broad critical assessment of such agents and studies remains inconclusive. Reassuringly, while there are no validated instruments that assess adherence and engagement of patients using conversational agents, the research community is however utilizing some measures, though they are not currently universally agreed upon. Going forward, studies should aim to include assessments of therapeutic efficacy. For example, for

depression, the “Severity Measure for Depression—Adult” is adapted from the Patient Health Questionnaire 9 (PHQ-9) and can be used to monitor treatment progress. Finally, to the best of our knowledge, no studies have been conducted regarding the interest among psychiatrists in conversational agents. Strikingly, many psychiatrists considered it unlikely that technology would ever be able to provide empathetic care as well as or better than the average psychiatrist.³⁴ Clinician engagement is necessary in order to integrate conversational agents into psychiatric practice and should be assessed with a modified engagement metric similar to those used for patients.

Unexplored Areas

Although conversational agent research is expanding, several areas remain understudied, primarily specific illness populations. Most research in the last year was conducted in adults, with the average participant age in studies as 34 years. An estimated 75% of all lifetime mental disorders emerge by age 24, and 50% emerge by age 14.³⁵ This highlights an important understudied period of intervention in which detection, monitoring, and treatment may have long-standing benefit on the trajectory of these young patient’s lives. While no studies assessing emergency response were discovered, there is emerging work on whether conversational agents are able to understand an emergency situation and appropriately respond.^{36,37}

Notably, 6 of the 7 conversational agents primarily had a text interface, and only 1 included a voice interface. By design, text is more discreet, which may allow patients to feel more comfortable using conversational agents in public. This is particularly true when sharing personal information regarding their emotions. In 2013, the Pew Research Center reported that of the U.S. adults who use digital voice assistants, 60% cite that they use digital voice assistants because “spoken language feels more natural than typing.”¹⁴ To our knowledge, a direct text to voice comparison of acceptance and therapeutic efficacy has not yet been conducted. More research into which modality is more effective in certain psychiatric conditions is needed.

These findings align with other reviews, concluding that while conversational agent interventions for mental health problems are promising, more robust experimental design is needed. In a review by Gaffney et al., metrics for engagement and reporting were inconsistent.³⁸ Another review by Bibault et al. focusing on oncology patients suggests that the scarcity of clinical trials in evaluating conversational agents contrasts with the increasing number of patients poised to use them.

It is important to characterize the use case of conversational agents. At present, conversational agents can potentially augment, but not replace, clinical care. To this end, conversational agents may best serve as a means of increasing access to care, such as to see a clinician. They could provide lists of mental health clinicians in the area or

recommend patients speak to their primary care physician regarding specific concerns that the conversational agent may not be capable of handling.

Further, while privacy and security remain a major concern in the use of technology in health-care settings, the sensitive nature of mental health may present a greater risk to patients. Little has been done to understand what steps, if any, are taken by commercially available conversational agents and whether sensitive, private, and vulnerable information shared by patients about their mental health is sufficiently safeguarded.

Limitations

It is important to note that there remain several limitations in evaluating conversational agents in mental health. Our search term aimed to be inclusive but may have missed some crucial studies, especially as research may be published in journals outside of those with a health-care focus chosen for this literature search. Our search term did not include terms such as “voice assistant,” “smart assistant,” and “dialog system” (these were also not included in our prior 2018 study), and these terms may have identified further studies. While comparing research changes between 2018 and 2020 offers useful insight, the high degree of heterogeneity between studies in this space continues to limit direct comparison. As highlighted in our prior review, the heterogeneity of reporting metrics continues to prevent the drawing of firm conclusions around already-limited use cases, and without a standardized metrics reporting framework, these limitations may persist further.

Conclusion

Conversational agents have continued to gain interest across the public health and global health research communities. This review revealed few, but generally positive, outcomes regarding conversational agents’ diagnostic quality, therapeutic efficacy, or acceptability. Despite the increase in research activity, there remains a lack of standard measures for evaluating conversational agents in regard to these qualities. We recommend that standardization of conversational agent studies should include patient adherence and engagement, therapeutic efficacy, and clinician perspectives. Given that patients are able to access a wide range of conversational agents on their mobile devices at any time, clinicians must carefully consider the quality and efficacy of these options, given such heterogeneity of available data.

Authors’ Note

Aditya Nrusimha Vaidyam and Danny Linggonegoro contributed equally to this work.

Acknowledgment

John Torous reports unrelated research support from Otsuka outside the scope of this work.

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) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: National Institute of Mental Health (grant ID: 1K23MH116130-01).

ORCID iDs

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

John Torous  <https://orcid.org/0000-0002-5362-7937>

References

- Wainberg ML, Scorza P, Shultz JM, et al. Challenges and opportunities in global mental health: a research-to-practice perspective. *Curr Psychiatry Rep.* 2017;19(5):28. doi:10.1007/s11920-017-0780-z.
- Depression [accessed 2020 Mar 28]. 2020. <https://www.who.int/news-room/fact-sheets/detail/depression>.
- Dieleman JL, Baral R, Birger M, et al. US spending on personal health care and public health, 1996-2013. *JAMA.* 2016; 316(24):2627-2646. doi:10.1001/jama.2016.16885.
- Flowers L, Houser A, Noel-Miller C, et al. Medicare spends more on socially isolated older adults. *AARP Public Policy Institute*; 2017. doi:10.26419/ppi.00016.001.
- Torous J, Myrick KJ, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. *JMIR Ment Health.* 2020;7(3): e18848. doi:10.2196/18848.
- Smith AC, Thomas E, Snoswell CL, et al. Telehealth for global emergencies: implications for coronavirus disease 2019 (COVID-19). *J Telemed Telecare.* Published online Mar 20, 2020. doi:10.1177/1357633X20916567.
- Pew Research Center. Voice assistants used by 46% of Americans, mostly on smartphones [accessed 2020 Mar 28]. 2020. <https://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/>.
- National Public Media. The smart audio report [accessed 2020 Mar 28]. 2020. <https://www.nationalpublicmedia.com/insights/reports/smart-audio-report/>.
- Feldman MJ, Hoffer EP, Barnett GO, Kim RJ, Famiglietti KT, Chueh HC. Impact of a computer-based diagnostic decision support tool on the differential diagnoses of medicine residents. *J Grad Med Educ.* 2012;4(2):227-231. doi:10.4300/JGME-D-11-00180.1.
- 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. doi:10.1093/jamia/ocy072.
- Miner AS, Milstein A, Hancock JT. Talking to machines about personal mental health problems. *JAMA.* 2017;318(13): 1217-1218. doi:10.1001/jama.2017.14151.
- Miner AS, Shah N, Bullock KD, Arnow BA, Bailenson J, Hancock J. Key considerations for incorporating conversational AI in psychotherapy. *Front Psychiatry.* 2019;10:746. doi: 10.3389/fpsy.2019.00746.
- Bendig E, Erb B, Schulze-Thuesing L, Baumeister H. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health—a scoping review. *VER.* Published online August 20, 2019:1-13. doi:10.1159/000501812.
- Philip P, Micoulaud-Franchi JA, 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(1):1-7. doi:10.1038/srep42656.
- Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *Can J Psychiatry.* 2019;64(7):456-464. doi:10.1177/0706743719828977.
- Adekunle L, Chen R, Morrison L, et al. Association between financial links to indoor tanning industry and conclusions of published studies on indoor tanning: systematic review. *BMJ.* 2020;368. doi:10.1136/bmj.m7.
- Bickmore TW, Trinh H, Olafsson S, et al. Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and google assistant. *J Med Internet Res.* 2018;20(9):e11510. doi:10.2196/11510.
- Bickmore TW, Puskar K, Schlenk EA, Pfeifer LM, Sereika SM. Maintaining reality: relational agents for antipsychotic medication adherence. *Interact Comput.* 2010;22(4):276-288. doi:10.1016/j.intcom.2010.02.001.
- Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. *JMIR Ment Health.* 2018;5(4):e64. doi:10.2196/mental.9782.
- Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR Mhealth Uhealth.* 2018;6(11):e12106. doi:10.2196/12106.
- Jungmann SM, Klan T, Kuhn S, Jungmann F. Accuracy of a chatbot (Ada) in the diagnosis of mental disorders: comparative case study with lay and expert users. *JMIR Form Res.* 2019;3(4):e13863. doi:10.2196/13863.
- Martínez-Miranda J, Martínez A, Ramos R, et al. Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour. *J Med Syst.* 2019;43(8):246. doi:10.1007/s10916-019-1387-1.
- Philip P, Dupuy L, Auriacombe M, et al. Trust and acceptance of a virtual psychiatric interview between embodied conversational agents and outpatients. *NPJ Digit Med.* 2020;3(1):1-7. doi:10.1038/s41746-019-0213-y.
- Provoost S, Ruwaard J, van Breda W, Riper H, Bosse T. Validating automated sentiment analysis of online cognitive behavioral therapy patient texts: an exploratory study. *Front Psychol.* 2019;10. doi:10.3389/fpsyg.2019.01065.
- Suganuma S, Sakamoto D, Shimoyama H. An embodied conversational agent for unguided internet-based cognitive behavior therapy in preventative mental health: feasibility and

- acceptability pilot trial. *JMIR Ment Health*. 2018;5(3):e10454. doi:10.2196/10454.
26. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Ment Health*. 2017;4(2):e19. doi:10.2196/mental.7785.
 27. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, O'Donnell J. Response to a relational agent by hospital patients with depressive symptoms. *Interact Comput*. 2010;22(4):289-298. doi:10.1016/j.intcom.2009.12.001.
 28. 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;4. doi:10.3389/frobt.2017.00051.
 29. Tielman ML, Neerincx MA, Bidarra R, Kybartas B, Brinkman WP. 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. doi:10.1007/s10916-017-0771-y.
 30. 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;10:39-46. doi:10.1016/j.invent.2017.10.002.
 31. Tielman ML, Neerincx MA, van Meggelen M, Franken I, Brinkman W-P. How should a virtual agent present psychoeducation? Influence of verbal and textual presentation on adherence. *Technol Health Care*. 2017;25(6):1081-1096. doi:10.3233/THC-170899.
 32. Shinozaki T, Yamamoto Y, Tsuruta S. Context-based counselor agent for software development ecosystem. *Computing*. 2015;97(1):3-28. doi:10.1007/s00607-013-0352-y.
 33. 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. doi:10.1016/j.pec.2017.04.015.
 34. Doraiswamy PM, Blease C, Bodner K. Artificial intelligence and the future of psychiatry: insights from a global physician survey. *Artif Intell Med*. 2020;102:101753. doi:10.1016/j.artmed.2019.101753.
 35. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Arch Gen Psychiatry*. 2005;62(6):593-602. doi:10.1001/archpsyc.62.6.593.
 36. Miner AS, Milstein A, Schueller S., Hegde R., Mangurian C., Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Intern Med*. 2016;176(5):619-625. doi:10.1001/jamainternmed.2016.0400.
 37. Kocaballi AB, Quiroz JC, Rezazadegan D, et al. Responses of conversational agents to health and lifestyle prompts: investigation of appropriateness and presentation structures. *J Med Internet Res*. 2020;22(2):e15823. doi:10.2196/15823.
 38. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. *JMIR Ment Health*. 2019;6(10). doi:10.2196/14166.