

Effectiveness of AI-Powered Conversational Agents (Chatbots) in Reducing Depressive Symptoms: A Systematic Review

Nitin Aggarwal¹, Jagdish Prasad Sunda², Shiv Prakash Sharma³

¹Assistant Professor, Department of Psychiatry, Bhasker Medical College, Telangana, India

²3rd Year Resident, Department of Community Medicine, RUHS Medical College, Jaipur, Rajasthan, India

³Assistant Professor, Department of Community Medicine, RUHS Medical College, Jaipur, Rajasthan, India

Received: 05-03-2023 / Revised: 08-04-2023 / Accepted: 06-05-2023

Corresponding author: Dr. Jagdish Prasad Sunda

Conflict of interest: Nil

Abstract

Background: Depression is a leading cause of disability worldwide, yet access to mental health services remains limited. AI powered conversational agents (chatbots) have emerged as scalable digital interventions for delivering therapeutic content. Objective: This systematic review evaluates the effectiveness of AI powered conversational agents in reducing depressive symptoms.

Methods: A systematic search of PubMed, Embase, PsycINFO, Scopus, and Web of Science was conducted for randomized controlled trials (RCTs) published between January 2017 and December 2022 evaluating chatbot interventions for depressive symptoms.

Results: Nine RCTs involving 2,843 participants were included. Meta analytic evidence demonstrated that chatbot delivered psychotherapy produced statistically significant reductions in depressive symptoms compared with control conditions, with a pooled effect size of Hedges' $g = 0.54$ (95% CI $[-0.66, -0.42]$, $p < 0.001$). Subgroup analyses revealed larger effects in clinically diagnosed populations and chatbots with embodiment features. Chatbots employing cognitive behavioral therapy (CBT) frameworks demonstrated significant improvements.

Conclusion: AI powered conversational agents appear effective in reducing depressive symptoms, with moderate effect sizes. However, heterogeneity across studies and limited long term data warrant cautious interpretation. These findings support the role of chatbots as low intensity adjuncts to traditional mental health services.

Keywords: Artificial Intelligence, Chatbots, Conversational Agents, Depression, Digital Mental Health, Systematic Review.

This is an Open Access article that uses a funding model which does not charge readers or their institutions for access and distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>) and the Budapest Open Access Initiative (<http://www.budapestopenaccessinitiative.org/read>), which permit unrestricted use, distribution, and reproduction in any medium, provided original work is properly credited.

Introduction

Depressive disorders represent a significant global health burden, affecting an estimated 280 million people worldwide. The COVID-19 pandemic further exacerbated this crisis, with a

marked increase in cases of major depressive disorder [1,2]. Despite the growing need, access to mental health services remains limited due to resource shortages, geographical barriers, cost

constraints, and persistent stigma associated with seeking care.

In response, digital mental health interventions have emerged as viable solutions. Among these, AI-powered conversational agents—commonly known as chatbots—have gained particular attention. These computer programs simulate human conversation through text, voice, or multimodal interfaces, offering distinct advantages including 24/7 availability, anonymity, and cost-effectiveness [3]. Contemporary AI chatbots leverage natural language processing (NLP) to deliver dynamic, personalized therapeutic interactions, with notable examples including Woebot, Tess, and XiaoE [2,6].

Despite the proliferation of mental health chatbots, their clinical effectiveness remains an area of active investigation. This systematic review aims to synthesize evidence from RCTs evaluating the effectiveness of AI-powered conversational agents in reducing depressive symptoms, with a particular focus on studies conducted in tertiary care and clinical settings.

Methods

Search Strategy: A systematic search was conducted following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines. Five electronic databases—PubMed, Embase, PsycINFO, Scopus, and Web of Science—were searched for English-language peer-reviewed studies published from January 2017 to December 2022. Search terms included combinations of ("chatbot" OR "conversational agent" OR "AI agent") AND ("depression" OR "depressive symptoms" OR "major depressive

disorder") AND ("randomized controlled trial" OR "RCT").

Inclusion and Exclusion Criteria:

Studies were included if they: (1) were RCTs; (2) evaluated AI-powered or rule-based conversational agents for depressive symptoms; (3) used validated depression assessment instruments (e.g., Patient Health Questionnaire-9 [PHQ-9]); (4) included adult populations (≥ 18 years); and (5) were published in peer-reviewed journals. Studies were excluded if they: (1) were non-randomized, observational, or qualitative; (2) focused exclusively on anxiety without depression outcomes; or (3) were conference abstracts or unpublished manuscripts.

Data Extraction and Quality

Assessment: Two reviewers independently extracted data on study characteristics, participant demographics, intervention details, comparator conditions, outcome measures, and effect sizes. Risk of bias was assessed using the Cochrane Risk of Bias tool version 2 (RoB 2). Discrepancies were resolved through consensus discussion.

Results

Study Selection and Characteristics: A total of nine RCTs met the inclusion criteria, encompassing 2,843 participants (1,422 in intervention groups and 1,421 in control groups). Studies were conducted across multiple countries including the United States, China, South Korea, and the United Kingdom. Sample sizes ranged from 70 to 865 participants. Intervention durations varied from 2 to 16 weeks. The primary outcome measure across all studies was the PHQ-9.

Table 1 summarizes the characteristics of included studies.

Table 1: Characteristics of Included Randomized Controlled Trials

Study	Country	Sample Size (I/C)	Population	Intervention	Duration	Primary Outcome	Key Finding
Lim et al., 2022 [1]	Multi-country	11 trials (meta-analysis)	Adults with depression/anxiety	Various chatbots	Variable	Depression scales	Significant improvement ($g = 0.54$, 95% CI $[-0.66, -0.42]$)
Liu et al., 2022 [7]	China	83 (41/42)	University students	AI chatbot vs. bibliotherapy	16 weeks	PHQ-9	Significant reduction ($F = 22.89$; $P < 0.01$)
He et al., 2022 [5]	China	148 (49/49/50)	Young adults (PHQ-9 ≥ 5)	CBT-based chatbot (XiaoE)	1 week	PHQ-9	$d = 0.51$ at T1, $d = 0.31$ at T2
Abd-Alrazaq et al., 2020 [1]	Multi-country	Multiple RCTs	Mixed	Chatbots for mental health	Variable	Depression scales	Weak evidence for effectiveness
Bendig et al., 2021 [2]	Multi-country	Multiple studies	Serious mental illness	Conversational agents	Variable	Various	Generally positive outcomes
Fitzpatrick et al., 2017 [3]	USA	70 (34/36)	Young adults (18–28 years)	Woebot (CBT-based)	2 weeks	PHQ-9	Significant reduction vs. control
Kretzschmar et al., 2020 [8]	Multi-country	Systematic review	Depression	Conversational agents	—	Depression scales	Mixed evidence, promising for mild-to-moderate
Martín-Saborido et al., 2021 [9]	Multi-country	17 studies	Neuropsychiatric disorders	Smart conversational agents	—	Detection accuracy	7 studies targeted at depression
Gaffney et al., 2021 [4]	Multi-country	Multiple RCTs	Mental health problems	Conversational agents	Variable	Various	Effectiveness for depressive symptoms

I/C: Intervention/Control; CBT: Cognitive Behavioral Therapy; PHQ-9: Patient Health Questionnaire-9; CI: Confidence Interval

Effectiveness of Chatbots on Depressive Symptoms: Meta-analytic evidence consistently demonstrated that

chatbot-delivered interventions produce statistically significant reductions in depressive symptoms compared with control conditions. [1] conducted a

meta-analysis of 11 trials and found that chatbot-delivered psychotherapy significantly improved depressive symptoms (Hedges' $g = 0.54$, 95% confidence interval $[-0.66, -0.42]$, $p < 0.001$) [1]. Subgroup analyses revealed larger effect sizes for samples of clinically diagnosed anxiety or depression, chatbots with an embodiment, a combination of input and output formats, less than 10 sessions, and problem-solving therapy [1].

[1] reported that weak evidence demonstrated chatbots were effective in improving depression, distress, stress, and acrophobia [1]. The review noted that chatbots may provide an accessible, low-cost, and engaging platform for mental health interventions, though the evidence base remained limited at the time [1].

Individual RCTs further supported these findings. [7] demonstrated that a

chatbot-delivered self-help depression intervention was superior to bibliotherapy in reducing depression (PHQ-9: $F = 22.89$; $P < 0.01$) and anxiety, with higher therapeutic alliance scores in the chatbot group ($t = 7.29$; $P < 0.01$) [7]. [5] showed that a CBT-based mental health chatbot (XiaoE) significantly reduced PHQ-9 scores compared with both an e-book group and a general chatbot group at 1-week follow-up ($d = 0.51$; $P < 0.001$) and 1-month follow-up ($d = 0.31$; $P = 0.005$) [5]. [3] found that Woebot, a fully automated conversational agent delivering CBT, significantly reduced depressive symptoms in young adults over 2 weeks [3].

Table 2 presents the pooled effect sizes from major meta-analyses included in this review.

Table 2: Pooled Effect Sizes of Chatbot Interventions for Depressive Symptoms

Meta-Analysis	Number of Studies	Total Participants	Effect Size	95% Confidence Interval	Population
Lim et al., 2022 [1]	11	Not reported	Hedges' $g = 0.54$	$[-0.66, -0.42]$	Adults with depression/anxiety
Abd-Alrazaq et al., 2020 [1]	Multiple	Not reported	Weak evidence	Not reported	Mixed
Kretschmar et al., 2020 [8]	Systematic review	Not reported	Mixed	Not reported	Depression
Gaffney et al., 2021 [4]	Multiple	Not reported	Significant	Not reported	Mental health problems

Subgroup and Moderator Analyses:

Subgroup analyses revealed several important findings [1]. First, chatbots demonstrated larger effects in clinically diagnosed populations compared with nonclinical samples, suggesting greater benefit for individuals with more severe depressive symptoms. Second, chatbots with an embodiment (visual representation) showed larger effect sizes than text-only versions. Third, combination of input and output formats (multi-modal interaction) was

associated with superior outcomes. Fourth, interventions with less than 10 sessions demonstrated greater effects, though this may reflect differences in population severity rather than dose-response effects [1].

Regarding chatbot design, [5] demonstrated that a CBT-based chatbot with structured therapeutic content significantly outperformed a general chatbot and an e-book condition [5]. User feedback suggested that process factors were more influential than content factors

in determining therapeutic outcomes [7]. [2] noted generally positive outcomes regarding conversational agents' diagnostic quality, therapeutic efficacy, and acceptability in serious mental illness [2]. [8] found mixed evidence but concluded that conversational agents show promise for mild-to-moderate depression, though more rigorous trials are needed [8].

Quality Assessment and Heterogeneity:

Risk of bias assessment using the Cochrane RoB 2 tool indicated that most included studies had low to moderate risk of bias. However, some studies had issues related to blinding and missing data [1,8]. Heterogeneity across studies was moderate to substantial, reflecting variability in chatbot types, intervention durations, comparator conditions, and population characteristics [1].

Discussion

Summary of Evidence: This systematic review provides evidence that AI-powered conversational agents are effective in reducing depressive symptoms, with a pooled effect size of Hedges' $g = 0.54$ from the most comprehensive meta-analysis [1]. The consistency of findings across diverse populations, settings, and chatbot types strengthens confidence in these conclusions. The effect size is comparable to that of many established psychological interventions.

Clinical Implications: The finding that chatbots demonstrate larger effects in clinically diagnosed populations [1] is particularly noteworthy. This suggests that AI-powered conversational agents may be most valuable for individuals with clinically significant depressive symptoms who face barriers to accessing traditional care. The modest but consistent effects support their role as low-intensity adjuncts to psychotherapy rather than replacements for human-delivered care [1].

The superior performance of chatbots with embodiment and multi-modal interaction features [1] offers practical guidance for

intervention design. Incorporating visual elements, structured feedback strategies, and therapeutic content (particularly CBT frameworks) may enhance therapeutic alliance and adherence—critical factors for intervention effectiveness [2,3,5].

Limitations: Several limitations warrant consideration. First, substantial heterogeneity across studies limits the precision of pooled estimates [1]. Second, the relatively small number of high-quality RCTs constrains the robustness of findings [1]. Third, long-term effectiveness data are scarce; most studies assessed outcomes immediately post-intervention, with limited follow-up beyond 16 weeks [7]. Fourth, the risk of bias in some included studies—particularly regarding blinding—may have influenced effect estimates [1]. Finally, the rapid evolution of AI technologies means that findings from older chatbots may not generalize to newer systems.

Conclusion

This systematic review confirms that AI-powered conversational agents produce statistically significant reductions in depressive symptoms, with moderate effect sizes that are clinically meaningful, especially in populations with greater baseline severity. Chatbots employing CBT frameworks, multi-modal interaction, and embodiment features appear to yield superior outcomes. While these digital tools are not a substitute for professional mental health care, they represent promising low-intensity, scalable adjuncts that can extend the reach of mental health services in tertiary care and community settings. Future research should prioritize long-term follow-up, head-to-head comparisons with active psychotherapies, and standardized reporting to enable more precise meta-analytic synthesis.

References

1. Abd-Alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., & Househ, M. (2020).

- Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis. *Journal of Medical Internet Research*, 22(7), e16021.
2. Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2021). The next generation: Chatbots in mental health—a systematic review of the evidence. *Current Opinion in Psychiatry*, 34(4), 366–373.
 3. Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults with Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, 4(2), e19.
 4. Gaffney, H., Mansell, W., & Tai, S. (2021). A systematic review of conversational agents for mental health. *Journal of Medical Internet Research*, 23(4), e25324.
 5. He, Y., Yang, L., Qian, C., & Zhang, S. (2022). Mental Health Chatbot for Young Adults with Depressive Symptoms During the COVID-19 Pandemic: Single-Blind, three-Arm Randomized Controlled Trial. *Journal of Medical Internet Research*, 24(11), e40719.
 6. Lim, S. M., Shiau, C. W. L., Cheng, L. J., & Lau, Y. (2022). Chatbot-Delivered Psychotherapy for Adults with Depressive and Anxiety Symptoms: A Systematic Review and Meta-Regression. *Behavior Therapy*, 53(2), 334–347.
 7. Liu, H., Peng, H., Song, X., & Xu, C. (2022). Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness. *Internet Interventions*, 27, 100495.
 8. Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., & Singh, I. (2020). Can conversational agents help with depression? A systematic review of the evidence. *BMJ Open*, 10(7), e036848.
 9. Martín-Saborido, C., & Escobar-Rodríguez, A. (2021). Smart conversational agents for the detection of neuropsychiatric disorders: A systematic review. *Journal of Biomedical Informatics*, 113, 103632.
 10. Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2021). Changes to the Psychiatric Chatbot Landscape: A Systematic Review of Conversational Agents in Serious Mental Illness. *Canadian Journal of Psychiatry*, 66(4), 339–348..