Proceedings of the 5th International Conference on International Law and Legal Policy
DOI: 10.54254/2753-7048/2026.37410

Advances in Research on the Effectiveness of Artificial Intelligence-Based Cognitive Behavioral Therapy Chatbots for Alleviating Anxiety and Depression Symptoms: A Literature Review

Qingyang She

School of Health and Social Sciences, James Cook University, Singapore, Singapore
qingyang.she@my.jcu.edu.au

Abstract. In recent years, advances in artificial intelligence (AI) have accelerated the development of digital mental health services, and cognitive behavioral therapy (CBT)-based chatbots have become an increasingly studied approach for anxiety and depression. These systems can deliver structured psychological support with relatively low cost, broad accessibility, continuous availability, and immediate interaction. This literature review synthesizes evidence from randomized controlled trials (RCTs), systematic reviews, and meta-analyses evaluating AI CBT chatbots for anxiety and depression. Current evidence suggests that these interventions can produce short-term reductions in mild-to-moderate depressive and anxiety symptoms, although effects are generally small to moderate and are more consistent for depression. However, confidence in these findings is limited by small or selective samples, short intervention and follow-up periods, substantial variation in chatbot design and control conditions, and limited evidence of long-term clinical benefit. Important concerns also remain regarding privacy, data governance, cultural and algorithmic bias, crisis recognition, and the reliability of AI-generated responses. Overall, AI CBT chatbots appear most defensible as accessible self-help or adjunctive tools rather than replacements for therapist-led CBT. Future research should prioritize larger and more diverse clinical samples, active comparison conditions, longer follow-up, transparent safety reporting, and carefully supervised human-AI models of care.

Keywords: Artificial intelligence, chatbot, cognitive behavioral therapy, anxiety, depression

1. Introduction

Anxiety and depression are major contributors to the global burden of mental ill-health and can substantially impair daily functioning, relationships, work, and quality of life [1]. The scale of this burden has increased demand for psychological support, while access to timely and evidence-based care remains uneven. Improving the reach and efficiency of mental health services has therefore become an important clinical and research priority.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2026.37410

Cognitive behavioral therapy (CBT) is one of the most established psychological treatments for anxiety and depression. CBT proposes that emotional distress is shaped partly by patterns of thinking and behavior, and treatment aims to identify maladaptive cognitions and modify behavioral responses through structured techniques [2]. Clinical guidelines continue to recommend CBT as an evidence-based treatment for depressive disorders [3]. Yet therapist-led CBT can be difficult to access because of treatment costs, waiting times, shortages of trained professionals, and geographic inequalities in service provision [4]. These barriers have encouraged interest in digital approaches that can extend psychological support beyond conventional clinical settings.

Digital mental health has developed rapidly alongside advances in natural language processing (NLP) and, more recently, large language models (LLMs). AI chatbots can deliver elements of CBT through conversational interfaces, including psychoeducation, mood monitoring, cognitive restructuring, behavioral activation, and problem-solving [5]. Systems such as Woebot, Wysa, and Youper illustrate how structured psychological content can be delivered through mobile or web-based interaction. Their continuous availability, anonymity, relatively low delivery cost, and scalability may make them useful where conventional services are difficult to access [4].

Evidence for these tools has grown, but the literature remains methodologically mixed. RCTs and meta-analyses report reductions in depressive and anxiety symptoms in some settings, while the magnitude and consistency of effects vary across studies [4, 6-8]. Chatbots differ in therapeutic content, degree of automation, conversational flexibility, and use of AI, while studies also differ in sample characteristics, outcome measures, intervention length, and control conditions. These differences make direct comparison difficult and weaken conclusions about sustained clinical efficacy. The move toward generative AI introduces further questions concerning privacy, bias, clinical safety, and responsibility for inappropriate or inaccurate responses.

This review therefore evaluates current evidence on AI-powered CBT chatbots for anxiety and depression. It first outlines the CBT principles that underpin these tools, then examines evidence from RCTs, systematic reviews, and meta-analyses. It subsequently compares AI-supported intervention with therapist-led CBT and considers major ethical, methodological, and practical limitations. Particular attention is given to whether current evidence justifies AI CBT chatbots as stand-alone treatment or supports a more limited role within self-help, stepped-care, and human-AI collaborative models.

2. Theoretical foundations of AI-powered CBT chatbots

CBT is a structured, evidence-based psychological approach widely used in the treatment of depression and anxiety [2, 3]. Its central model proposes that interpretations of events influence emotional and behavioral responses. Persistent negative automatic thoughts, cognitive biases, and maladaptive behavioral patterns can therefore maintain psychological distress. CBT targets these processes through techniques such as cognitive restructuring, behavioral activation, self-monitoring, and problem-solving, helping individuals develop more balanced interpretations and more adaptive patterns of behavior.

AI CBT chatbots translate selected components of this structured approach into digital interactions. Their main practical advantage is not that they reproduce a full course of psychotherapy, but that they can deliver repeatable CBT exercises at scale and with few scheduling barriers. An anonymous interface may also reduce concerns about stigma and make initial engagement easier for some users [4]. These features support a plausible role for chatbots in low-intensity self-help, early support, or between-session practice, particularly for people with mild-to-moderate symptoms.

DOI: 10.54254/2753-7048/2026.37410

At the same time, digitizing CBT techniques does not reproduce all mechanisms of therapist-led treatment. Clinical judgment, case formulation, flexible adaptation, risk assessment, and the therapeutic alliance depend on contextual understanding that current AI systems cannot reliably provide [5]. The theoretical rationale for AI CBT is therefore strongest when the technology is viewed as a delivery channel for selected evidence-based techniques rather than as an autonomous therapist. This distinction is important when interpreting outcome studies and determining the appropriate clinical role of these systems.

3. Effects of AI cognitive behavioral therapy chatbots on anxiety and depression symptoms

Existing RCTs, systematic reviews, and meta-analyses suggest that AI-based conversational interventions can reduce depressive and anxiety symptoms in the short term, although reported effects are generally small to moderate and vary across studies. Li et al. [7] found that AI-based conversational agents produced beneficial effects on several mental health outcomes, including depression and psychological distress. Zhong et al. [9] similarly reported short-course benefits for depressive and anxiety symptoms. Farzan et al. [4] concluded that CBT-oriented chatbots show potential as accessible psychological support, while also emphasizing limitations in study quality, intervention duration, and follow-up. Taken together, these findings support cautious optimism rather than evidence of equivalence to established psychotherapy.

The Woebot RCT by Fitzpatrick et al. [6] remains an influential early study. Seventy young adults with self-reported symptoms of anxiety and depression were randomized to two weeks of Woebot or an information-only control. Depressive symptoms declined more in the Woebot condition, whereas anxiety decreased in both groups without a clear advantage for the chatbot. The study demonstrated that structured CBT content could be delivered through an automated conversational interface and produce measurable short-term change. However, its small sample, brief intervention, and young non-clinical population limit generalizability. It also illustrates why symptom improvement should not automatically be attributed to AI itself: the observed benefit may reflect CBT content, repeated self-monitoring, engagement, or other nonspecific aspects of receiving digital support.

Evidence from more recent studies also requires careful interpretation. Wang et al. [10], for example, evaluated a CBT-based AI chatbot among Chinese university students and reported improvement in depression and loneliness. This study is relevant because it extends the evidence base beyond Western samples and highlights the importance of contextual and cultural factors. However, it should not be treated as direct evidence of an anxiety outcome because anxiety was not the principal outcome reported in the cited study. More broadly, differences between depression and anxiety findings across the literature may partly reflect intervention design: behavioral activation and structured thought-monitoring can be readily translated into digital tasks, whereas anxiety treatment may require exposure work and situation-specific clinical decisions that are harder to automate safely.

Positive findings for anxiety should nevertheless not be dismissed. Reviews including systems such as Wysa report reductions in both anxiety and depressive symptoms in some user groups [4]. Possible mechanisms include frequent self-monitoring, immediate prompts, repeated CBT practice, and easier access to support. Yet these mechanisms are difficult to separate from nonspecific effects such as psychoeducation, expectations of benefit, or simply receiving regular attention. Studies that rely on wait-list, information-only, or no-intervention controls cannot establish whether improvement is caused by AI functionality, CBT techniques, engagement with a digital tool, or a combination of these factors.

DOI: 10.54254/2753-7048/2026.37410

Methodological heterogeneity further complicates interpretation. The term "AI chatbot" covers rule-based conversational systems, NLP-based tools, and newer generative models, which differ substantially in flexibility, personalization, and safety characteristics [5]. Samples range from university students and general community users to people with chronic illness or clinically significant symptoms. Outcomes are also measured with different self-report instruments, including the Patient Health Questionnaire-9 (PHQ-9), Generalized Anxiety Disorder-7 (GAD-7), and Depression Anxiety Stress Scales-21 (DASS-21). A statistically significant reduction on these scales indicates symptom change but does not necessarily demonstrate remission, functional recovery, or reduced need for professional treatment. Short follow-up periods also leave uncertainty about maintenance and relapse.

The current evidence therefore supports a limited but clinically relevant conclusion. AI CBT chatbots can provide short-term symptom support, especially for mild-to-moderate depression, while evidence for anxiety is positive but less consistent. The literature does not establish that these systems can replace comprehensive therapist-led CBT. Their most defensible roles are low-intensity self-help, support while waiting for care, and structured practice alongside professional treatment. Stronger conclusions will require larger clinical samples, active control conditions, clearer reporting of chatbot functions, and longer follow-up.

4. Comparison between AI-powered CBT chatbots and traditional CBT

AI CBT chatbots and therapist-led CBT share techniques such as cognitive restructuring, behavioral activation, self-monitoring, and problem-solving, but they differ in treatment intensity and clinical capability. Therapist-led CBT involves assessment, case formulation, individualized planning, ongoing review, and adaptation to the patient's response. A therapist can also integrate verbal and non-verbal information, assess comorbidity and risk, and respond to unexpected changes in presentation. Clinical guidance emphasizes structured delivery by appropriately trained practitioners when CBT is used as a formal treatment [3].

AI chatbots offer a different set of strengths. They can provide structured exercises without an appointment, operate continuously, and reduce some geographic and financial barriers to support. Li et al. [7] found beneficial mental health outcomes across AI-based conversational interventions, supporting their potential as scalable low-intensity tools. However, the comparison with traditional CBT must remain cautious. Much of the chatbot literature uses wait-list, information-only, or no-intervention controls rather than a full course of therapist-led CBT. Evidence that a chatbot performs better than limited support is therefore not evidence that it performs as well as professional psychotherapy.

The therapeutic relationship also marks an important difference. In therapist-led CBT, collaboration, shared goals, trust, and the therapeutic bond can shape engagement and treatment outcomes. A chatbot can simulate empathic language, but this response is generated by programmed rules or statistical language models rather than by clinical understanding of the person's experience. Ambiguous statements, culturally specific meanings, changes in risk, and complex comorbidity may therefore be misunderstood or handled too generally. This limitation becomes especially important when users interpret fluent conversational responses as evidence of professional competence.

For these reasons, current evidence supports a stepped-care or collaborative role for AI CBT chatbots more strongly than a replacement model. People with mild-to-moderate symptoms may use chatbots for self-monitoring, psychoeducation, skills practice, or support between treatment contacts. People with severe symptoms, substantial functional impairment, complex comorbidity, or safety concerns require professional assessment and therapist-led care. In a human-AI model,

DOI: 10.54254/2753-7048/2026.37410

automated systems could manage repetitive and structured tasks while clinicians retain responsibility for formulation, risk assessment, treatment decisions, and complex therapeutic work.

5. Challenges and future directions for AI cognitive behavioral therapy chatbots

The broader use of AI CBT chatbots raises ethical and clinical issues that extend beyond symptom efficacy. Mental health conversations can contain highly sensitive information about emotions, relationships, health history, and risk. Users may not understand how these data are stored, processed, shared, or used to improve a system. Meaningful consent therefore requires clear and accessible information about data use, retention, and third-party access. Data minimization and strong governance are particularly important because a privacy failure in a mental health context may create both direct harm and loss of trust in care.

Reliability is another concern, particularly as conversational systems move from rule-based programs toward generative AI. Rule-based systems are limited in flexibility but their outputs are comparatively predictable. Generative models can produce more natural and personalized responses, yet they can also generate inaccurate, inconsistent, or contextually inappropriate content [5]. In mental health settings, this is not merely a usability problem. Failure to recognize deterioration, risk, or the need for urgent professional support can have clinical consequences. Systems therefore require explicit boundaries, tested escalation procedures, and clear communication that automated responses do not constitute professional diagnosis or crisis care.

Bias and cultural fit also require greater attention. Chatbots trained or evaluated mainly in particular languages, age groups, or cultural settings may not interpret distress, family relationships, stigma, or help-seeking in the same way across populations. The evidence base itself is concentrated in relatively young and digitally engaged samples, which limits confidence in generalization to older adults, people with low digital literacy, and more diverse clinical populations. Future evaluation should therefore examine not only average symptom change but also who benefits, who disengages, and whether performance differs across demographic and cultural groups.

Future research should also improve the quality of comparison and safety evidence. Larger RCTs, active control groups, longer follow-up, and transparent reporting of adverse events are needed to distinguish specific therapeutic effects from general engagement effects. Studies should describe whether a system is rule-based, NLP-based, or generative, because these technologies have different capabilities and risks. A promising direction is supervised human-AI collaboration: chatbots may support mood monitoring, psychoeducation, reminders, and structured CBT exercises, while clinicians retain responsibility for assessment, individualized formulation, risk management, and treatment decisions. Such a model preserves the accessibility of digital support without assuming that conversational fluency is equivalent to clinical competence.

6. Conclusion

This review examined whether AI-powered CBT chatbots can reduce symptoms of anxiety and depression and how their role should be understood in relation to established psychological care. Across RCTs, systematic reviews, and meta-analyses, the evidence indicates modest short-term benefits, with findings generally more consistent for depressive symptoms than for anxiety. The strongest support is for low-intensity use among people with mild-to-moderate symptoms rather than for treatment of complex or high-risk clinical presentations.

The evidence also remains less definitive than the rapid growth of these technologies may suggest. Small and selective samples, brief interventions, heterogeneous chatbot designs, weak

DOI: 10.54254/2753-7048/2026.37410

comparison conditions, and limited follow-up restrict conclusions about long-term effectiveness. In addition, privacy, bias, cultural fit, unreliable generated content, and limitations in risk recognition create clinical and ethical concerns that symptom outcome studies alone cannot resolve.

AI CBT chatbots should therefore be understood as potential additions to mental health care rather than substitutes for trained clinicians. Their most appropriate current uses include accessible self-help, support while waiting for services, and structured practice alongside professional treatment. Future research should test these roles in larger and more diverse clinical populations, use stronger control conditions and longer follow-up, and evaluate safety as carefully as efficacy. Human-AI collaborative models may offer the most credible path forward by combining the accessibility of digital tools with professional judgment, accountability, and individualized care.

References

[1] World Health Organization. (2025, September 30). Mental disorders. *World Health Organization*.

[2] Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*. International Universities Press.

[3] National Institute for Health and Care Excellence. (2022). *Depression in adults: Treatment and management (NICE Guideline NG222)*. NICE.

[4] Farzan, M., Ebrahimi, H., Pourali, M., & Sabeti, F. (2025). Artificial intelligence-powered cognitive behavioral therapy chatbots: A systematic review. *Iranian Journal of Psychiatry, 20*(1), 102–110. https://doi.org/10.18502/ijps.v20i1.17395

[5] Im, C.-H., & Woo, M. (2025). Clinical efficacy, therapeutic mechanisms, and implementation features of cognitive behavioral therapy–based chatbots for depression and anxiety: Narrative review. *JMIR Mental Health, 12*, e78340. https://doi.org/10.2196/78340

[6] 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. https://doi.org/10.2196/mental.7785

[7] Li, H., Zhang, R., Lee, Y.-C., Kraut, R. E., & Mohr, D. C. (2023). Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *npj Digital Medicine, 6*, Article 236. https://doi.org/10.1038/s41746-023-00979-5

[8] Sohn, J.-S., Ha, B.-G., Park, S., Kim, J.-J., Lee, E., Oh, H., Lee, S., & Kim, E. (2026). Systematic review and meta analysis of chatbots in the management of depressive and anxiety symptoms. *npj Digital Medicine, 9*, Article 377. https://doi.org/10.1038/s41746-026-02566-w

[9] Zhong, W., Luo, J., & Zhang, H. (2024). The therapeutic effectiveness of artificial intelligence-based chatbots in alleviation of depressive and anxiety symptoms in short-course treatments: A systematic review and meta-analysis. *Journal of Affective Disorders, 356*, 459–469. https://doi.org/10.1016/j.jad.2024.04.057

[10] Wang, Y., Li, X., Zhang, Q., Yeung, D., & Wu, Y. (2025). Effect of a cognitive behavioral therapy–based AI chatbot on depression and loneliness in Chinese university students: Randomized controlled trial with financial stress moderation. *JMIR mHealth and uHealth, 13*, e63806. https://doi.org/10.2196/63806