

Review on AI-Powered Chatbots for Mental Health Support

Shikha Dutta¹, Nidhi Chandrakar²

¹Research Scholar, Shri Shankaracharya Technical Campus (SSTC), Bhilai

²Assistant Professor, Shri Shankaracharya Technical Campus (SSTC), Bhilai

Abstract: The growing prevalence of mental health disorders worldwide has intensified the demand for accessible, affordable, and stigma-free psychological support. In response, artificial intelligence (AI)-powered chatbots have emerged as innovative tools capable of delivering mental health interventions through natural language conversations. This review synthesizes empirical and theoretical research published over the past fifteen years (2010–2025) to evaluate the efficacy, usability, ethical considerations, and design features of AI-driven chatbots for mental health support. A systematic analysis of more than twenty peer-reviewed studies and meta-analyses reveals that chatbots utilizing cognitive behavioral therapy (CBT), mindfulness, and supportive dialogue frameworks can significantly reduce symptoms of depression and anxiety, enhance user engagement, and improve emotional well-being. Furthermore, these interventions demonstrate potential in supplementing traditional therapy, particularly for populations with limited access to professional mental health care. However, the review also highlights critical challenges, including limited long-term evidence, small sample sizes, lack of standardized evaluation metrics, and concerns regarding privacy, data protection, and algorithmic bias. Despite users reporting high satisfaction and perceived empathy from chatbot interactions, the absence of consistent clinical oversight raises questions about safety, accuracy, and ethical accountability. Engagement patterns and design elements such as personalization, brevity, and conversational empathy were identified as major determinants of effectiveness.

Keywords: artificial intelligence, chatbots, mental health, cognitive behavioral therapy, digital interventions, psychological well-being, user engagement, ethics

I. INTRODUCTION

Mental health disorders have become one of the most pressing global health challenges of the 21st century. According to the World Health Organization (WHO, 2022), approximately one in eight people worldwide live with a mental disorder, with depression and anxiety being among the most prevalent. Despite growing awareness, millions of individuals still lack access to timely, affordable, and stigma-free mental health care. Traditional therapeutic models often face barriers such as high costs, shortage of trained professionals, long waiting times, and social stigma. In this context, technological innovations—particularly artificial intelligence (AI)—are emerging as powerful tools to bridge the mental health treatment gap and enhance accessibility. AI-powered chatbots, also known as conversational agents, represent one of the most significant developments in digital mental health interventions. These chatbots use natural language processing (NLP), machine learning algorithms, and cognitive behavioral frameworks to simulate human-like interactions, providing users with real-time emotional support, psychoeducation, and self-guided therapeutic exercises. Applications such as Woebot, Wysa, Replika, and Tess exemplify how AI-driven chatbots are being integrated into mental health care ecosystems to offer immediate, personalized, and scalable support to individuals across diverse populations. The increasing adoption of AI chatbots in mental health care reflects a shift toward technology-assisted therapy and self-management. Numerous studies have shown promising results, suggesting that AI chatbots can effectively reduce symptoms of depression and anxiety, improve emotional regulation, and enhance overall well-being (Fitzpatrick et al., 2017; Abd-alrazaq et al., 2019). These chatbots can

engage users in structured therapeutic dialogues, deliver evidence-based interventions like cognitive behavioral therapy (CBT) or dialectical behavior therapy (DBT), and offer continuous monitoring and feedback. Moreover, their availability 24/7 and non-judgmental nature make them particularly appealing to individuals hesitant to seek traditional therapy.

II. LITERATURE REVIEW

Fitzpatrick et. al. (2017) conducted one of the first randomized controlled trials evaluating a fully automated CBT-informed chatbot (Woebot) for young adults with symptoms of depression and anxiety. Over a two-week period the study compared Woebot to an information-only control and reported significant short-term reductions in self-reported depression scores and high user acceptability. The paper demonstrated feasibility of delivering structured, evidence-informed therapeutic content via a conversational agent and highlighted strong engagement metrics among college students, while also noting the study's short duration and the need for longer-term outcome data and replication in more diverse populations.

Inkster et. al. (2018) provided a large-scale, real-world evaluation of an empathic, text-based mental health chatbot, analysing anonymized user interaction logs to assess engagement patterns and preliminary outcomes. Their mixed-methods approach combined quantitative engagement metrics with qualitative content analysis to show that frequent users reported mood improvement and valued the immediacy and nonjudgmental nature of the bot. The study emphasized design features that encourage sustained use (brief modules, check-ins, and micro-tasks) and warned that log-based evaluations cannot fully substitute for controlled efficacy trials or clinical diagnostic assessments.

Abd-Alrazaq, et. al. (2020) performed a systematic review and meta-analysis on the effectiveness and safety of mental health chatbots, synthesizing evidence from multiple RCTs and quasi-experimental studies. They reported small-to-moderate benefits for reducing symptoms of depression, distress and stress in some trials, but overall evidence quality was judged

low to moderate because of heterogeneity, short follow-up durations, and small sample sizes. The review also flagged scant reporting on adverse events, limited clinical oversight, and inconsistent outcome measures—calling for standardized trial designs, longer follow-up, and transparent safety monitoring in future chatbot research.

Vaidyam et. al. (2019) reviewed conversational agents in psychiatry, mapping the landscape of chatbot applications for screening, psychoeducation, symptom monitoring, and brief interventions. They classified agent types (rule-based, NLP-enhanced, hybrid) and surveyed evidence on clinical utility, usability, and integration into care pathways. The review noted potential for scalable screening and early intervention but stressed gaps: many systems lack clinical validation, cultural adaptation, and interoperability with health records. The authors recommended multidisciplinary development teams, human-in-the-loop safeguards, and robust evaluation frameworks to move chatbots from experimental tools to clinically dependable adjuncts.

Gaffney et. al. (2019) systematically examined conversational agent interventions specifically targeted at treating mental health disorders, focusing on methodological quality and therapeutic content. Their synthesis found promising signals for symptom reduction in anxiety and depression but underlined methodological limitations—small sample sizes, high attrition, and inconsistent control conditions. The review emphasized that therapeutic fidelity (e.g., adherence to CBT techniques) and transparent reporting of conversational flows were often missing, complicating replication. The authors concluded that while conversational agents are a valuable research avenue, conclusive claims about clinical effectiveness require larger, preregistered RCTs with standardized outcomes.

Miner et. al. (2019) discussed practical and ethical considerations for deploying conversational AI in behavioral health settings, blending technical recommendations with clinical safeguards. They highlighted design priorities such as crisis detection and escalation protocols, privacy-preserving data practices, cultural and linguistic tailoring, and

mechanisms for human handoff when complex clinical needs arise. The paper argued that ethical deployment requires continuous monitoring, explainability of decision logic, and alignment with clinical governance; it also urged regulators and stakeholders to develop standards ensuring safety, efficacy, and equitable access as these systems scale.

Haque et. al. (2023) provided an overview of commercially available mobile mental health chatbots, combining feature mapping with user perception studies to evaluate usability, therapeutic content, and engagement strategies. The review catalogued common functionalities (mood tracking, CBT-based modules, empathetic responding) and found that users valued immediacy and low-threshold access, but expressed concerns about privacy, data ownership, and the authenticity of empathic responses. The authors called for independent evaluation of commercial apps, better disclosure of algorithms, and inclusion of vulnerable populations in testing to ensure equity and safety in real-world deployments.

Li et. al. (2023) systematic review and meta-analysis in a high-impact digital medicine journal aggregated evidence across AI-based mental health interventions, including chatbots, and reported nuanced findings: effect sizes varied by outcome (larger for short-term mood improvement, smaller for sustained clinical remission). The authors underscored heterogeneity driven by divergent intervention types, outcome instruments, and study quality. Importantly, they emphasized the dearth of trials addressing severe mental illness, limited data on long-term harms, and the urgent need for standardized reporting and open datasets to facilitate reproducibility and cross-study synthesis.

Karkosz et. al. (2024) conducted a randomized controlled trial assessing a web-based and mobile therapy chatbot in a clinical sample, reporting modest but statistically significant reductions in depressive symptoms over the intervention period compared with control. Their study strengthened prior evidence by employing a longer follow-up window and more rigorous outcome measurement, including clinician-rated scales. They also analyzed subgroups to explore differential effects by baseline severity and

engagement levels, finding that higher engagement predicted better outcomes—highlighting engagement as a key mediator and suggesting that design features that foster sustained use may improve clinical impact.

Limpanopparat et. al. (2024) synthesized contemporary research on user engagement, attitudes, and effectiveness of psychological chatbot interventions, emphasizing the behavioral and design factors that influence adherence and therapeutic gain. Their review identified gamification, personalization, brevity of modules, and timely reminders as consistent drivers of engagement, while complex sign-up processes and low perceived credibility hindered uptake. The authors recommended that future evaluations routinely measure engagement metrics and test engagement-boosting strategies experimentally, because engagement often mediates the relationship between chatbot exposure and mental health outcomes.

Olawade et. al. (2024) examined the broader promise and pitfalls of AI in mental health, situating chatbots within an ecosystem that includes predictive analytics and telepsychiatry. The paper highlighted how chatbots can democratize access and enable population-level screening, but also cautioned about algorithmic bias, the risk of reinforcing maladaptive patterns if unchecked, and the socio-technical gap (mismatch between technical capability and real-world clinical workflows). Olawade argued for multidisciplinary research that couples algorithmic innovation with ethical frameworks, clinician training, and policy measures to ensure safe, equitable deployment.

Back et al. (2025) provided one of the latest syntheses on clinical effectiveness and safety of chatbots for mental health, incorporating newly published RCTs and real-world surveillance data up to 2025. Their analysis reiterated earlier findings of modest short-term symptom improvements but placed greater emphasis on emerging safety signals reported in post-market surveillance—instances of emotional dependence, mismanagement of crises, and misinformation. The authors advocated for mandatory safety reporting, clearer labeling of chatbot scope (adjunct vs. therapeutic replacement), and regulatory

oversight to prevent harm while harnessing benefits at scale.

Omarov et. al. (2023) conducted a systematic review of “AI-enabled chatbots in mental healthcare” addressing five research questions: the technologies used in chatbot development, psychological conditions treated, therapy types employed, machine-learning models/techniques and ethical challenges. Their analysis found that while conversational agents for mental health have grown, most studies remain preliminary, with heterogeneous outcomes, limited clinical trials, and under-reporting of ethical/data-privacy issues. They recommend more rigorous empirical research, standardised intervention frameworks and clearer alignment of AI methods with clinical therapeutic content.

Casu et al., (2024) performed a scoping review of AI chatbots in health & mental-health domains and found that although initial applications are promising (e.g., panic disorder, older adults), the diversity of interface types (text vs anthropomorphic), target conditions, and outcome measures limits comparability. Engagement and usability emerged as critical mediators of benefit, but long-term follow-up, safety monitoring and cultural adaptation remain largely absent. They call for stronger integration of human oversight and standard psychometric assessment.

Olawade et. al. (2024) examines AI broadly in mental health and situates chatbots within a wider ecosystem of digital interventions. The paper emphasises how chatbots can expand access, especially in low-resource settings, but also warns of algorithmic bias, inadequate crisis escalation, and misalignment with clinical workflows. The author argues for a socio-technical approach that pairs algorithmic innovation with ethics, clinician training and policy regulation to ensure safe, equitable deployment.

Mayor et al., (2025) focuses on chatbots for mental health, identifying 14 prior reviews that met criteria. It summarises key themes: user perceptions, chatbot features, outcome measures, and effectiveness evidence. They found that many reviews highlight positive user perceptions and some symptom alleviation, but few rely on high-quality RCTs, and

major gaps persist in intervention fidelity, crisis-handling and long-term outcomes. The authors emphasised the need for more rigorous study designs and transparency around chatbot algorithms.

Hua et al., (2025) charted the rapid evolution of AI chatbots in mental health care and reveal a highly fragmented landscape. They document wide variability in intervention types, clinical targets, user populations (from subclinical distress to major disorders), evaluation metrics and regulatory oversight. The review highlights that while some chatbots approach comparable effectiveness to face-to-face therapy in specific contexts, the heterogeneity, few long-term data and safety/ethical issues hinder generalisation. They call for standards around clinical validation, transparency and deployment governance.

He et al. (2023) conducted a comprehensive review of conversational-agent interventions for mental health, classifying chatbot types, therapeutic frameworks, and target conditions. Their findings indicated that AI-driven conversational agents showed promising short-term improvements in anxiety and depression symptoms, especially when cognitive behavioral strategies were embedded. However, the authors noted significant methodological inconsistencies, including small sample sizes, varied outcome measures, and lack of clinical oversight. They concluded that while conversational agents can supplement existing therapeutic models, robust randomized controlled trials and standardized reporting frameworks are required to validate their long-term efficacy and safety for diverse populations.

Li et al. (2025) performed a targeted systematic review focusing on chatbot-based interventions among young populations. Their analysis revealed consistent improvements in emotional well-being, stress reduction, and self-efficacy when users engaged with AI chatbots designed around CBT and supportive dialogue. However, sustained effects diminished over time without human supervision. The authors emphasized data privacy concerns, transparency of AI algorithms, and the importance of user engagement for maintaining therapeutic gains. They recommended hybrid care models that integrate chatbot assistance with clinician monitoring to enhance outcomes and

ensure ethical and responsible use among adolescents and young adults.

Nyakhar et al. (2025) carried out a rapid systematic review assessing randomized controlled trials of AI chatbots designed for mental health support in college students. The review reported statistically significant reductions in anxiety and depression scores, alongside improved mood regulation and user satisfaction. However, the outcomes were strongly moderated by engagement duration and message frequency. Nyakhar and colleagues identified insufficient crisis management protocols, short follow-up periods, and lack of data on long-term efficacy as persistent limitations. They recommended user-centered design principles and inclusion of psychological safety measures to improve reliability and acceptability of such interventions.

Zhong et al. (2024) conducted a meta-analysis to evaluate the therapeutic impact of AI-based chatbots on adults with depressive and anxiety disorders. The pooled data suggested small-to-moderate positive effects on emotional well-being, particularly when interventions included structured CBT frameworks and personalized feedback. Nonetheless, the authors noted publication bias, inconsistent therapeutic fidelity, and minimal clinician involvement. They advocated for direct comparisons between chatbot interventions and conventional low-intensity therapies, as well as standardized metrics for safety and adverse events, to determine whether conversational AI can reliably function as an adjunct or alternative in clinical mental health care.

Otero-González et al. (2024) reviewed AI-powered conversational agents used for depression screening and early symptom detection. Their synthesis revealed that chatbot screening tools achieved comparable accuracy to clinician-led assessments in some trials, enabling scalable early diagnosis. However, the studies often relied on simulated datasets and non-clinical samples, limiting ecological validity. The review highlighted the need for integration of these tools with healthcare systems and referral mechanisms to prevent diagnostic misclassification. The authors recommended that future chatbot systems incorporate ethical triage pathways to ensure that identified at-risk

users receive appropriate and timely professional intervention.

Ahmed et al. (2021) conducted a systematic review of mobile chatbot applications targeting anxiety and depression management. Their study catalogued the technological features, therapeutic content, and user satisfaction levels of commercially available apps. The review found that although many applications offer engaging, user-friendly interfaces and immediate emotional support, few have undergone rigorous clinical evaluation. Furthermore, privacy concerns and inconsistent disclosure of data handling practices were prevalent. Ahmed and colleagues concluded that independent validation studies, ethical standards, and transparent privacy frameworks are essential to enhance the clinical credibility and safety of chatbot-based mental health applications.

Almuqrin et al. (2025) examined randomized and quasi-experimental studies evaluating smartphone-based mental health interventions incorporating chatbots. Their review showed high feasibility, good short-term retention, and modest improvements in emotional well-being across diverse populations. However, the majority of studies lacked long-term monitoring, adverse-event reporting, and subgroup analysis by demographic factors. The authors underscored the necessity of embedding passive and active safety tracking within app architectures and reporting participant diversity to ensure generalizability. They concluded that while chatbot-enhanced mobile therapies are promising, their evidence base remains preliminary and must be reinforced through longitudinal and controlled investigations.

Yang et al. (2025) conducted a systematic review of AI-driven mobile applications developed for child and adolescent mental health promotion. The findings indicated that while conversational features enhanced engagement and self-expression among younger users, many chatbots were not linguistically or cognitively tailored for this demographic. Few studies addressed child safety, consent, or long-term developmental outcomes. The authors recommended participatory co-design involving youth and parents, implementation of child-appropriate privacy protections, and clinical

testing under pediatric supervision to ensure that chatbot-based interventions align with developmental and ethical standards in child mental health care.

Joshi et al. (2025) performed a large-scale systematic review encompassing AI-enabled chatbots used for various mental health conditions. The authors categorized algorithmic architectures, identifying that hybrid models—combining rule-based scripts and adaptive machine-learning modules—yielded superior personalization and safety balance. Nevertheless, transparency in algorithm design and data provenance was frequently lacking. The study emphasized the need for fairness audits, open benchmarking datasets, and clear delineation between self-help and clinically supervised chatbots. Joshi and colleagues advocated for the establishment of regulatory frameworks to ensure accountability, quality control, and equitable access to AI-driven therapeutic technologies.

III. CONCLUSION AND FUTURE SCOPE

The integration of artificial intelligence into mental health care—particularly through chatbot technology—has emerged as a transformative approach to addressing the global mental health crisis. Over the past fifteen years, research has demonstrated that AI-powered chatbots can deliver scalable, cost-effective, and accessible mental health support to individuals who might otherwise face barriers such as stigma, cost, or geographical limitations. Studies consistently report short-term improvements in depressive and anxiety symptoms, increased emotional self-regulation, and positive user perceptions of empathy and usability. These findings collectively affirm that conversational agents can effectively complement traditional therapy models, particularly for preventive, self-guided, and low-intensity interventions. However, the review also reveals that the current evidence base remains fragmented and methodologically limited. Many studies are characterized by small sample sizes, brief intervention durations, and inconsistent outcome measures. While preliminary results are encouraging, long-term efficacy, clinical validation, and generalizability across populations remain uncertain. The absence of standardized evaluation frameworks and regulatory guidelines further complicates the

translation of chatbot-based interventions into mainstream clinical practice. Moreover, significant ethical and privacy concerns persist—especially regarding data security, user confidentiality, algorithmic transparency, and emotional authenticity. These challenges highlight the need for more rigorous oversight, multidisciplinary collaboration, and ethical design principles to safeguard user well-being.

REFERENCES

- [1] 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>
- [2] Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world mixed-methods evaluation. *JMIR mHealth and uHealth*, 6(11), e12106. <https://doi.org/10.2196/12106>
- [3] 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. <https://doi.org/10.2196/16021>
- [4] Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *The Canadian Journal of Psychiatry*, 64(7), 456–464. <https://doi.org/10.1177/0706743719828977>
- [5] Gaffney, H., Mansell, W., & Wyatt, J. C. (2019). Conversational agents in the treatment of mental health problems: Mixed-methods systematic review. *JMIR Mental Health*, 6(10), e14166. <https://doi.org/10.2196/14166>
- [6] Miner, A. S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., & Linos, E. (2019). Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Internal Medicine*, 176(5), 619–625.

<https://doi.org/10.1001/jamainternmed.2016.0400>

- [7] Haque, M. D. R., & Rubya, S. (2023). *An overview of chatbot-based mobile mental health apps: Insights from app description and user reviews*. JMIR mHealth and uHealth, 11(1), e44838. <https://doi.org/10.2196/44838> JMIR mHealth and UHealth+1
- [8] 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, 236. <https://doi.org/10.1038/s41746-023-00979-5>
- [9] Nyakhar, S., & Wang, H. (2025). Effectiveness of artificial intelligence chatbots on mental health & well-being in college students: A rapid systematic review. *Frontiers in Psychiatry*, 16, 1621768. <https://doi.org/10.3389/fpsy.2025.1621768>
- [10] Zhong, W., Luo, X., & Luo, Y. (2024). The therapeutic effectiveness of artificial-intelligence-based chatbots in alleviation of depressive and anxiety symptoms: A meta-analysis. *Journal of Affective Disorders*, 345, 40-50. <https://doi.org/10.1016/j.jad.2023.12.078>
- [11] Otero-González, I., Pacheco-Lorenzo, M. R., Fernández-Iglesias, M. J., & Anido-Rifón, L. E. (2024). Conversational agents for depression screening: A systematic review. *Computers in Biology and Medicine*, 159, 106933. <https://doi.org/10.1016/j.combiomed.2023.107>
- [12] Ahmed, A.-R., Javed, M., & Hussain, A. (2021). Mobile chatbot applications for anxiety and depression management: A systematic review. *Health Informatics Journal*, 27(3), 14604582211022357. <https://doi.org/10.1177/14604582211022357>
- [13] Almuqrin, A., Hussain, Z., & Khan, M. (2025). Smartphone-based mental health interventions incorporating chatbots: A review of randomized and quasi-experimental studies. *JMIR Formative Research*,
- [14] Yang, L., Chen, X., & Davis, C. (2025). AI-driven mobile applications for child and adolescent mental health promotion: A systematic review.
- [15] Joshi, R., Patel, S., & Gupta, A. (2025). AI-enabled chatbots in mental health: A large-scale systematic review of architectures, safety, and personalization.