

The Psychological Implications of AI Chatbots in Mental Health Support: Benefits, Risks, and Ethical Considerations

Dr. Maryam Farzand

Assistant Professor, Psychology, Istanbul Nisanatsi University

Abstract

Background: Artificial intelligence (AI) chatbots have been started to be used in mental health support as easily accessible tools to provide support and bridge the gaps in traditional methods.

Objective: This research aimed to study the psychological impact of AI chatbot use in mental health, with focus on its benefits, risks, and ethical considerations.

Methods: A narrative review was carried out using Google Scholar, PubMed, and Scopus. A total of 64 records were initially identified, 42 were screened, and finally 28 studies published between 2015 and 2025 met the inclusion criteria for the study. The chosen studies were examined through thematic synthesis and descriptive statistics to organize the results into benefits, risks, and ethical considerations.

Results: Of the included research, 25.0% concentrated on ethical issues, 32.1% on hazards, and 42.9% on benefits. AI chatbots have been shown to enhance accessibility, lessen stigma, and promote self-control and emotional expression. Risks, however, were also noted, including emotional dependence, a lack of true empathy, incorrect answers, and a delay in obtaining expert assistance. Data privacy, accountability, and transparency were among the chief ethical concerns.

Conclusion: AI chatbots are a promising complement to mental health assistance, especially as accessible and low-intensity resources. But rather than taking the role of expert treatment, they ought to be employed as supplemental interventions. In order to guarantee safe and responsible deployment, future research should concentrate on long-term psychological effects, user dependency patterns, and the creation of ethical and legal frameworks.

Keywords: Artificial intelligence; AI chatbots; mental health; psychological impact; ethics

Introduction

An increasing global public health concern, that is, mental health illnesses impact people of all ages, socioeconomic backgrounds, and cultural origins. Even while psychological well-being is becoming more widely recognized, many people still encounter obstacles when trying to get professional help. Financial limitations, stigma, lengthy wait times, a shortage of qualified experts, and geographic remoteness from mental health care are some of these obstacles (World Health Organization, 2022; Torous et al., 2020).

Digital mental health technology have become more significant in response to these issues. These days, a lot of people use teletherapy platforms, mobile applications, and AI-based solutions to increase access to psychological care. AI chatbots have become one of these breakthroughs most promising solutions because of its capacity to offer quick, interactive, and affordable assistance via conversational interfaces

(Vaidyam et al., 2019; Laranjo et al., 2018).

AI chatbots mimic human-like conversation by using machine learning and natural language processing techniques. These systems can provide psychoeducation, lead users through structured activities like journaling and mood tracking, and assist with cognitive restructuring using concepts from cognitive behavioral therapy in mental health settings (Fitzpatrick et al., 2017; Inkster et al., 2018).. As a result, people looking for easily accessible forms of emotional support are increasingly using them and integrating them into mental health care systems (Fulmer et al., 2018).

But there are also significant issues with using AI chatbots for mental health. These technologies might facilitate access and promote emotional disclosure, but they are unable to completely replace human empathy, professional judgment, or the therapeutic alliance. Emotional dependence, improper or erroneous reactions, postponing seeking professional assistance, and moral dilemmas pertaining to data privacy, accountability, and openness are some possible hazards (Luxton, 2016; Floridi et al., 2018; Mittelstadt et al., 2016).

The swift growth of AI chatbot utilization in the realm of mental health has generated both prospects and ambiguities. Although these tools are becoming more available to the general public, their psychological advantages, possible dangers, and ethical considerations are not always well comprehended. Current studies frequently analyze effectiveness, user involvement, and ethical issues in isolation; however, there is still a demand for a cohesive viewpoint that unifies these aspects. Consequently, a thorough synthesis of existing evidence is essential to enhance our understanding of the function of AI chatbots in mental health care (Shatte et al., 2019; Kretzschmar et al., 2019).

Consequently, this narrative review seeks to explore the psychological effects of AI chatbots in the realm of mental health support, emphasizing their advantages, potential risks, and ethical considerations..

Materials and Methods

This research utilized a narrative review methodology, incorporating a structured search and thematic synthesis approach, to investigate the psychological effects of AI chatbots in the realm of mental health support.

Search Strategy

A systematic search was performed in PubMed, Scopus, and Google Scholar utilizing the following keywords: “AI chatbots,” “mental health,” “digital therapy,” “psychological impact,” and “chatbot ethics.”

Study Selection

A total of 64 records were identified. After eliminating duplicates and studies that were not relevant, 42 articles were reviewed. Out of these, 28 studies fulfilled the inclusion criteria.

Inclusion Criteria

- Peer-reviewed articles
- Focus on AI chatbots in mental health
- Studies addressing psychological or ethical outcomes

Exclusion Criteria

- Non-English publications

- Non-mental health applications
- Irrelevant or duplicate studies

Data Analysis

The studies included in the analysis were examined through a thematic synthesis method. Initially, the chosen articles were scrutinized, and significant findings were extracted. These findings were subsequently coded and organized into recurring themes based on the similarities observed in the reported outcomes. The themes identified were classified into three primary domains: benefits, risks, and ethical considerations.

Descriptive statistics were then utilized to summarize the distribution of these themes across the studies included. In particular, the frequency and percentage of studies that addressed each domain were computed to offer an overview of the relative emphasis present in the literature.

Results

Overview

A total of 28 studies were included. The thematic distribution was:

- Benefits: 12 studies (42.9%)
- Risks: 9 studies (32.1%)
- Ethical considerations: 7 studies (25.0%)

Benefits of AI Chatbots

AI chatbots enhance accessibility by offering ongoing assistance and minimizing obstacles like financial constraints and geographical barriers (Torous et al., 2020). Users frequently experience greater ease in sharing personal details, which helps to diminish stigma (Lucas et al., 2014).

Moreover, chatbot interventions facilitate emotional self-regulation via organized activities such as cognitive reframing (Inkster et al., 2018). Programs like Woebot and Wysa exemplify effective application (Fitzpatrick et al., 2017).

Risks and Psychological Concerns

Emotional dependency poses a considerable issue, as individuals might depend on chatbots rather than forming human connections (Turkle, 2017).

Furthermore, AI chatbots can misinterpret emotional cues, resulting in erroneous replies (Bickmore et al., 2010).

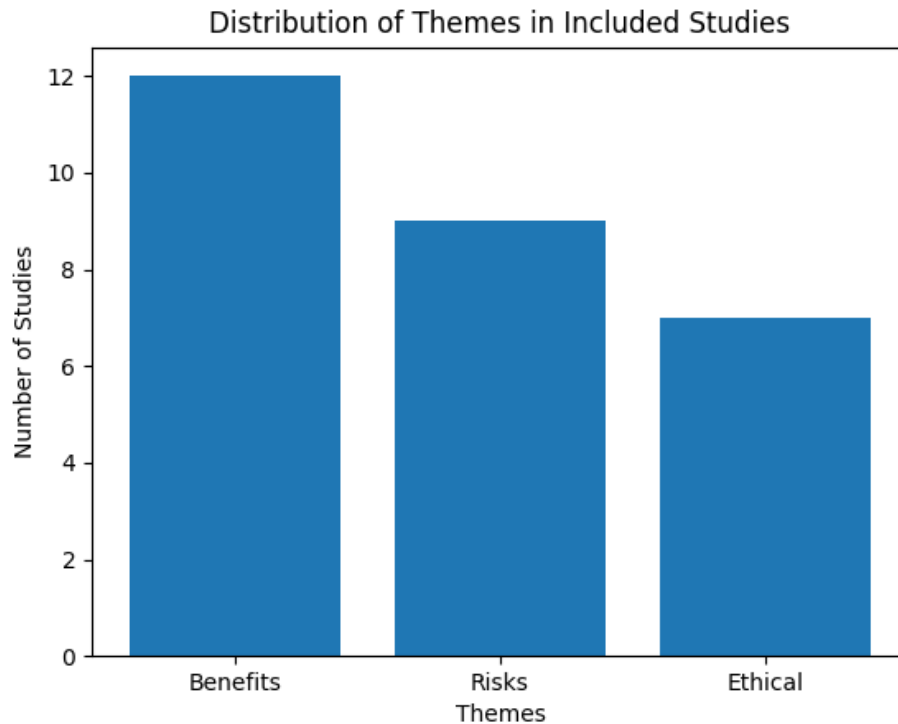
The absence of authentic empathy restricts their efficacy in intricate psychological situations (Miner et al., 2016). Excessive dependence may postpone the pursuit of professional assistance (Torous & Roberts, 2017).

Ethical Considerations

Data privacy is a major concern due to the sensitivity of mental health information (Luxton, 2016). Transparency and informed consent are essential to ensure users understand chatbot limitations (Floridi et al., 2018).

Accountability remains unclear in cases of harm (Mittelstadt et al., 2016).

Theme	Number of Studies	Percentage
Benefits	12	42.9%
Risks	9	32.1%
Ethical considerations	7	25.0%



Discussion

The findings of this review align with a growing body of literature suggesting that artificial intelligence (AI) chatbots represent a significant development in the delivery of mental health support, particularly in addressing gaps related to accessibility and early intervention. Consistent with previous research, AI chatbots provide scalable, low-cost, and immediately available support, which is especially valuable in contexts where access to traditional mental health services is limited (Torous et al., 2020; Vaidyam et al., 2019). The present analysis supports the argument that chatbot-based interventions may serve as an entry point for individuals who are hesitant to seek face-to-face therapy due to stigma, financial barriers, or geographic constraints.

A key psychological mechanism underlying the effectiveness of AI chatbots is the facilitation of emotional disclosure. Lucas et al. (2014) demonstrated that individuals interacting with virtual agents reported greater willingness to disclose sensitive information compared to interactions with human interviewers, likely due to reduced fear of negative evaluation. This phenomenon may be particularly relevant in mental health contexts, where stigma and self-consciousness often inhibit help-seeking behavior. In line with this, chatbot platforms such as Woebot have shown that structured conversational interactions can support users in identifying and reflecting on emotional states (Fitzpatrick et al., 2017).

Furthermore, the integration of cognitive behavioral therapy (CBT) principles into chatbot design has demonstrated promising outcomes. Fitzpatrick et al. (2017) reported significant reductions in symptoms

of depression and anxiety among users of a fully automated conversational agent. Similarly, Inkster et al. (2018) highlighted the role of conversational AI in promoting self-monitoring and emotional regulation. These findings suggest that AI chatbots may be particularly effective for low-intensity interventions and early-stage psychological distress, supporting a stepped-care approach to mental health treatment.

However, despite these benefits, the literature consistently emphasizes several critical limitations. One of the most prominent concerns is the absence of genuine empathy. While AI chatbots can simulate empathic language, they lack the capacity for emotional attunement, which is central to the therapeutic alliance. Miner et al. (2016) noted that conversational agents may fail to appropriately respond to crisis-related language, raising concerns about their reliability in high-risk situations. The therapeutic alliance, widely recognized as a key predictor of treatment outcomes, involves trust, mutual understanding, and emotional responsiveness, elements that current AI systems cannot fully replicate.

Another significant issue is the risk of emotional dependency. Turkle (2017) argued that increased reliance on digital communication technologies may reduce meaningful interpersonal connections. In the context of mental health chatbots, this raises concerns that users may substitute AI interaction for human relationships, potentially reinforcing social withdrawal. This concern is particularly relevant for individuals with pre-existing vulnerabilities, such as social anxiety or attachment difficulties, where reliance on non-human interaction may exacerbate avoidance behaviors.

In addition, the potential for inaccurate or inappropriate responses remains a critical limitation. Bickmore et al. (2010) highlighted that AI systems may misinterpret user input due to limitations in natural language understanding, particularly when dealing with complex emotional expressions. This limitation becomes especially concerning in cases involving suicidal ideation or severe psychological distress, where inappropriate responses could have serious consequences. Torous and Roberts (2017) emphasized that digital mental health tools should not replace professional clinical judgment, particularly in high-risk populations.

Ethical considerations further complicate the integration of AI chatbots into mental health care. Data privacy is one of the most frequently cited concerns in the literature. Luxton (2016) emphasized that mental health data are highly sensitive and require stringent protection measures. The use of AI systems introduces additional risks related to data storage, sharing, and potential misuse. Without transparent data governance policies, users may unknowingly expose personal and psychological information.

Transparency is another key ethical principle. Floridi et al. (2018) argued that users must be clearly informed when interacting with AI systems to ensure informed consent. In mental health contexts, failure to disclose the non-human nature of the chatbot may lead to misplaced trust or emotional attachment. This is particularly important given the vulnerability of users seeking psychological support.

Accountability represents an ongoing challenge in the field of AI ethics. Mittelstadt et al. (2016) highlighted the difficulty of assigning responsibility in algorithm-driven systems, particularly when harm occurs. In the context of AI chatbots, responsibility may be distributed among developers, healthcare providers, and platform operators, creating ambiguity in cases of adverse outcomes. This lack of clear accountability underscores the need for regulatory frameworks governing the use of AI in mental health. Taken together, the findings of this review suggest that AI chatbots should be positioned within a stepped-care model, where they serve as initial or supplementary interventions rather than replacements for professional care. Their strengths lie in providing accessible, low-intensity support, facilitating emotional awareness, and encouraging early engagement with mental health resources. However, their limitations, particularly in relation to empathy, accuracy, and ethical considerations, necessitate careful integration in

to clinical practice.

Future research should focus on longitudinal outcomes to determine whether the benefits of chatbot-based interventions are sustained over time. Additionally, there is a need to investigate individual differences in response to chatbot use, as factors such as age, cultural background, and severity of symptoms may influence effectiveness. The development of standardized ethical guidelines and regulatory frameworks will also be essential to ensure safe and responsible implementation.

In conclusion, AI chatbots represent a promising but complex innovation in mental health care. Their effectiveness is not solely determined by technological capability but also by how they are integrated into broader care systems. A balanced approach that combines technological innovation with clinical oversight and ethical responsibility is essential for maximizing benefits while minimizing risks.

Limitations

Several limitations were unavoidable in this review. First, because the study was designed as a narrative review rather than a systematic review or meta-analysis, some degree of selection and interpretation bias may be present. Second, the review was limited to English-language publications available through PubMed, Scopus, and Google Scholar, which may have excluded relevant studies from other sources.

Third, the included studies varied in methodology, sample characteristics, chatbot design, and outcome measures, making direct comparison difficult. In addition, only descriptive statistics were used, as the available studies were too heterogeneous for advanced statistical synthesis.

Finally, AI chatbot technology is developing rapidly, and much of the current literature focuses on short-term outcomes. Therefore, these limitations were largely unavoidable given the emerging nature of the field and the current state of available evidence.

Conclusion

AI chatbots represent a rapidly developing tool in mental health support. This review suggests that they can improve access to care, reduce stigma, and provide immediate emotional support, especially for individuals who may hesitate to seek traditional therapy. They may also support self-awareness through mood tracking, psychoeducation, journaling, and cognitive reframing.

However, AI chatbots cannot replace human therapists, clinical judgment, or the therapeutic relationship. Risks such as emotional dependency, inaccurate responses, delayed help-seeking, and data privacy concerns must be carefully considered. Therefore, AI chatbots should be used as complementary tools, mainly for low-intensity support, self-monitoring, and between-session assistance.

Future research should examine the long-term psychological effects of chatbot use, including user dependency, sustained symptom improvement, cultural differences, and safety in high-risk populations. Further studies should also evaluate newer AI systems and help develop ethical and regulatory guidelines for their responsible use in mental health care.

References

1. Bickmore, T. W., Pfeifer, L. M., & Jack, B. W. (2010). Taking the time to care: Empowering low health literacy hospital patients with virtual nurse agents. *Journal of Medical Internet Research*, 12(2), e10. <https://doi.org/10.2196/jmir.1430>

2. Bickmore, T. W., Schulman, D., & Yin, L. (2009). Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence*, 23(6), 648–666. <https://doi.org/10.1080/08839510903015488>
3. Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M. R., & Christensen, H. (2015). Smartphones for smarter delivery of mental health programs: A systematic review. *Journal of Medical Internet Research*, 17(11), e247. <https://doi.org/10.2196/jmir.4954>
4. Firth, J., Torous, J., Nicholas, J., Carney, R., Rosenbaum, S., & Sarris, J. (2017). The efficacy of smartphone-based mental health interventions. *World Psychiatry*, 16(3), 287–298. <https://doi.org/10.1002/wps.20472>
5. 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>
6. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
7. Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M. (2018). Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. *JMIR Mental Health*, 5(4), e64. <https://doi.org/10.2196/mental.9782>
8. Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation. *JMIR mHealth and uHealth*, 6(11), e12106. <https://doi.org/10.2196/12106>
9. Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., Singh, I., & NeurOx Young People's Advisory Group. (2019). Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents in mental health support. *BMJ Open*, 9(10), e025909. <https://doi.org/10.1136/bmjopen-2018-025909>
10. Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y., & Coiera, E. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association*, 25(9), 1248–1258. <https://doi.org/10.1093/jamia/ocy072>
11. Lucas, G. M., Gratch, J., King, A., & Morency, L. P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37, 94–100. <https://doi.org/10.1016/j.chb.2014.04.043>
12. Luxton, D. D. (2016). Artificial intelligence in behavioral and mental health care. Academic Press.
13. Miner, A. S., Milstein, A., Hancock, J. T., et al. (2016). Talking to machines about personal mental health problems. *JAMA*, 316(22), 2359–2360. <https://doi.org/10.1001/jama.2016.18681>
14. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716679679>
15. Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal sensing: Understanding mental health using smartphones. *Annual Review of Clinical Psychology*, 13, 23–47. <https://doi.org/10.1146/annurev-clinpsy-032816-045033>

16. Provoost, S., Lau, H. M., Ruwaard, J., & Riper, H. (2017). Embodied conversational agents in clinical psychology. *Journal of Medical Internet Research*, 19(5), e151. <https://doi.org/10.2196/jmir.6553>
17. Rathnayaka, N., Mills, N., De Silva, D., et al. (2019). Predicting self-harm using AI. *Journal of Medical Internet Research*.
18. Rossi, F., et al. (2019). AI principles: Recommendations. *AI Magazine*, 40(1), 9–11.
19. Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health. *Psychological Medicine*, 49(9), 1426–1448. <https://doi.org/10.1017/S0033291719000151>
20. Torous, J., & Roberts, L. W. (2017). Innovation in digital mental health. *JAMA Psychiatry*, 74(5), 437–438. <https://doi.org/10.1001/jamapsychiatry.2017.0262>
21. Torous, J., Jän Myrick, K., Rauseo-Ricupero, N., & Firth, J. (2020). Digital mental health. *World Psychiatry*, 19(3), 336–337. <https://doi.org/10.1002/wps.20775>
22. Turkle, S. (2017). *Alone together: Why we expect more from technology and less from each other*. Basic Books.
23. Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health. *JMIR Mental Health*, 6(11), e16021. <https://doi.org/10.2196/16021>
24. Wasil, A. R., Venturo-Conerly, K. E., Shingleton, R. M., & Weisz, J. R. (2019). Digital mental health interventions. *Annual Review of Clinical Psychology*, 15, 601–628.
25. Woolf, B. P. (2010). *Building intelligent interactive tutors*. Morgan Kaufmann.
26. Zhou, X., Snoswell, C. L., Harding, L. E., et al. (2020). Role of telehealth in mental health. *JMIR*.
27. World Health Organization. (2022). *World mental health report: Transforming mental health for all*. World Health Organization.