STUDIA DOCTORALIA
PSYCHOLOGY AND EDUCATIONAL SCIENCE
UNIVERSITY OF BUCHAREST

EDITORIAL

„In Therapy" with Algorithms

Denisa Berceanu

University of Bucharest

https://doi.org/10.47040/sdpsych.v16/2.186

In an evolving world where artificial intelligence increasingly reshapes our daily lives, and people are starting to talk more often to algorithms, one question inevitably emerges: is it safe and ethical to use AI for mental health care? Let's talk about that.

Artificial intelligence is already entering the therapeutic space through AI-driven chatbots designed to support people experiencing depression, anxiety, and stress. Some of the most studied tools, such as Woebot and Wysa, show encouraging results. For example, a randomized controlled trial (Suharwardy et al., 2023) evaluated postpartum women using Woebot for six weeks. The Woebot group showed significantly reduction in depression scores from baseline on one assessment tool that is commonly used in clinical practice. Over 70% achieved clinically significant improvement, and no adverse events were reported. Engagement was high, with participants interacting with the chatbot about five times per week. Similarly, Wysa demonstrated strong engagement among users with no or moderate anxiety symptoms (Chang, et al., 2024).

Moreover, according to some studies, chatbots can build therapeutic bonds comparable to that found in traditional face-to-face CBT, even within just a few days of use, suggesting that AI chatbots can successfully establish meaningful therapeutic engagement with users (Darcy et al., 2021). Mobile mental health apps are particularly promising for young people, who appreciate flexibility, cultural sensitivity, and personalized interfaces (Chen et al., 2025).

Diagnostic and monitoring applications of AI have shown promise as well. According to Cruz-González et al. (2025) AI tools can detect and predict mental health conditions with high accuracy using algorithms like support vector machines and random forests. Also, tools such as machine learning and natural language processing can analyze patient behavior and text for early intervention opportunities (Le Glaz et al., 2021). Yet, these tools often rely on narrow datasets, such as social media posts, which may not reflect diverse populations or contexts.

However, significant limitations persist. Chatbots and AI companions lack the empathy and contextual judgment necessary to form a true therapeutic alliance, something central to effective psychotherapy. Moreover, only 16% of LLM studies underwent clinical efficacy testing, with most (77%) still in early validation (Hua et al., 2025). The World Health Organization has emphasized that AI must not replace human clinicians, whose capacity for moral reasoning and emotional connection remains irreplaceable (WHO, 2021).

These tools offer 24/7, nonjudgmental support, especially for those facing stigma or lacking access to human therapists. However, in traditional psychotherapy, sessions are deliberately spaced to encourage reflection, integration, and the development of coping autonomy. By contrast, the on-demand nature of AI companions might foster emotional dependence rather than resilience if not carefully guided. Thus, while high engagement reflects user trust and accessibility, it also highlights the need for balanced use that promotes self-efficacy rather than reliance.

Also, evidence on AI-based conversational agents shows mixed outcomes. While some studies report meaningful reductions in depression and distress, others find no significant impact on overall psychological well-being (Li et al., 2023). Importantly, few studies assess long-term safety. Abd-Alrazaq et al. (2020) found only weak evidence supporting chatbot

efficacy and limited safety evaluations. In addition, generative AI has failed in real-world settings to detect suicidal ideation, sometimes giving inappropriate or harmful advice (Hua et al., 2025).

Bias is another threat. AI systems trained on non-representative data may produce skewed outputs that reinforce health disparities (Saeidnia et al., 2024). And while machine learning shows promise in predicting suicide risk (Kirtley et al., 2022), its real-world impact is constrained by data quality, ethical considerations, and the lack of scalable interventions for those identified as high-risk. As a result, while chatbots hold promise for expanding access to mental health support, current research is largely exploratory and lacks the robust clinical trials and interactional depth needed to validate their effectiveness (Mayor, 2025).

To address these concerns, the WHO has outlined six guiding principles for AI in health, including the protection of human autonomy, transparency, equity, and sustainability (WHO, 2021). These principles are not optional, they are the ethical backbone of responsible AI deployment in mental health care.

So, is it okay to use AI in therapy? Yes, if we remember who it serves. AI can guide and even comfort, but it cannot feel. It cannot truly sit with silence, read between the words, or hold space for pain. Let AI be the map, but let the human therapist remain the compass. Only together can they chart a path toward compassionate, ethical, and accessible care.

REFERENCES

Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & 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. doi: 10.2196/16021

Chang, C. L., Sinha, C., Roy, M., & Wong, J. C. M. (2024). AI-led mental health support (Wysa) for health care workers during COVID-19: service evaluation. *JMIR Formative Research, 8*, e51858. doi: 10.2196/51858

Chen, X., Chen, L., Song, D., Wang, Y., & Ryoo, J. (2025). Digital mental health for young people: sustainably managing stress and anxiety. *Mental Health and Digital Technologies, 2*(3), 157-182. https://doi.org/10.1108/MHDT-01-2025-0009

Cruz-Gonzalez, P., He, A. W. J., Lam, E. P., Ng, I. M. C., Li, M. W., Hou, R., ... & Vidaña, D. I. S. (2025). Artificial intelligence in mental health care: a systematic review of diagnosis, monitoring, and intervention applications. *Psychological medicine, 55*, e18. doi:10.1017/S0033291724003295

Darcy, A., Daniels, J., Salinger, D., Wicks, P., & Robinson, A. (2021). Evidence of human-level bonds established with a digital conversational agent: cross-sectional, retrospective observational study. *JMIR Formative Research, 5*(5), e27868. doi: 10.2196/27868

Hua, Y., Siddals, S., Ma, Z., Galatzer-Levy, I., Xia, W., Hau, C., ... & Torous, J. (2025). Charting the evolution of artificial intelligence mental health chatbots from rule-based systems to large language models: a systematic review. *World Psychiatry, 24*(3), 383-394. https://doi.org/10.1002/wps.21352

Kirtley, O. J., van Mens, K., Hoogendoorn, M., Kapur, N., & de Beurs, D. (2022). Translating promise into practice: A review of machine learning in suicide research and prevention. The *Lancet Psychiatry, 9*(3), 243–252. https://doi.org/10.1016/S2215-0366(21)00254-6

Le Glaz, A., Haralambous, Y., Kim-Dufor, D. H., Lenca, P., Billot, R., Ryan, T. C., ... & Lemey, C. (2021). Machine learning and natural language processing in mental health: systematic review. *Journal of medical Internet research, 23*(5), e15708. doi: 10.2196/15708

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*(1), 236. https://doi.org/10.1038/s41746-023-00979-5

Mayor, E. (2025). Chatbots and mental health: a scoping review of reviews. *Current Psychology*, 1-22. https://doi.org/10.1007/s12144-025-08094-2

Saeidnia, H. R., Hashemi Fotami, S. G., Lund, B., & Ghiasi, N. (2024). Ethical Considerations in Artificial Intelligence Interventions for Mental Health and Well-Being: Ensuring Responsible Implementation and Impact. *Social Sciences, 13*(7), 381. https://doi.org/10.3390/socsci13070381

Suharwardy, S., Ramachandran, M., Leonard, S. A., Gunaseelan, A., Lyell, D. J., Darcy, A., ... & Judy, A. (2023). Feasibility and impact of a mental health chatbot on postpartum mental health: a randomized controlled trial. *AJOG Global Reports, 3*(3), 100165. https://doi.org/10.1016/j.xagr.2023.100165

World Health Organization. (2021). *Ethics and governance of artificial intelligence for health*. World Health Organization. https://www.who.int/publications/i/item/9789240029200

World Health Organization. (2021). *WHO issues first global report on Artificial Intelligence (AI) in health and six guiding principles for its design and use*. www.who.int/news/item/28-06-2021 - who-issues-first-global-report-on-ai-in-health-and-six-guiding-principles-for-its-design-and-use/