

AI-driven therapy: advancing care or compromising it?

Abstract

Virtual psychotherapy platforms and artificial intelligence (AI) are revolutionising mental health treatment by providing 24/7, affordable, and easily accessible help. These tools do, however, present serious issues with empathy, cultural sensitivity, ethical protections, diagnostic precision, and data security, notwithstanding their potential. This article examines the developing role of AI in psychotherapy, going into its clinical uses, technological foundations, and moral conundrums. Based on ethical responsibility and scientific rigour, the conversation highlights a careful and well-rounded approach to integrating AI into psychotherapy.

Keywords: artificial intelligence, psychotherapy, mental health, tele-therapy, ethics, data security

Volume 15 Issue 3 - 2025

Rushabh Shah,¹ Tanusha Guneta,² Sneh Babra,³ Nitin K Sethi⁴

¹Neuropsychology Intern, University of Delhi, India

²Neuropsychology Intern, Indira Gandhi National Open University (IGNOU), India

³Neuropsychology Intern, National Forensic Sciences University, India

⁴Department of Neurology, New York-Presbyterian Hospital, NY, and Pushpawati Singhania Research Institute, India

Correspondence: Nitin K Sethi, MD, MBBS, FAAN, Chairman Neurosciences and Senior Consultant, PSRI Hospital, New Delhi, India

Received: June 20, 2025 | **Published:** July 08, 2025

Introduction

Psychotherapy is a form of psychological intervention aimed at helping individuals overcome various mental health issues.¹ It is also recommended for those with organic brain and neurological disorders to manage disability alongside pharmacological and rehabilitative treatment. Traditionally, psychotherapy involves establishing a patient-therapist relationship (i.e., therapeutic alliance), making diagnoses and then providing treatment through various modalities and techniques such as cognitive-behavioural and psychodynamic psychotherapy. It has typically involved a single patient and a single clinician.^{2,3}

The COVID-19 pandemic brought the limitations of face-to-face therapy and the use of technology into sharp focus. For the Indian population, some of the commonly reported stressors by the masses during this unprecedented crisis included fear of one's health and well-being of family, sense of isolation due to quarantine measures, fear of job loss, economic difficulties, loss of usual social systems and overabundance of misinformation.⁴ Together, these stressors contributed to a massive psychosocial impact. Several if not all countries reported higher-than-usual levels of distress, anxiety, and depression.⁵

While it was non-negotiable that these concerns be addressed, the infectious nature of the virus and the barriers created by containment strategies led to closing down of the traditional face-to-face facilities. Thus, telecounselling, previously considered to be a peripheral service, had rapidly emerged as an important tool for providing mental health care during the pandemic.^{6,7} Tele-therapy or e-therapy encompasses the use of digital technology to provide clinical services such as assessment and treatment. These may include use of telephone, email, chats, televideo communication technology and virtual reality to offer services similar to face-to-face therapy. They are helpful in reaching even the most vulnerable sections of society or during times of crises or disaster given the easy accessibility, anonymity, confidentiality, making way out of hindrances of distance and stigma.⁶ This shift set the stage for Artificial Intelligence to play a transformative role in psychotherapy, offering new opportunities but also raising significant challenges.

AI in psychotherapy: technology and applications

More recent trends involve innovations and opportunities offered by Artificial Intelligence (AI) in the psychotherapeutic process. AI simply refers to machine-based intelligence that operates within and impacts its environment. Beg et al.¹ distinguishes between the various key AI technologies in use in the domain of psychotherapy:

- 1. Machine Learning (ML)** focuses on data-driven hypothesis development, using data and algorithms to build models that can predict and classify on their own;
- 2. Natural Language Processing (NLP)** is a field devoted to the computational processing and interpretation of human language, especially unstructured text;
- 3. Deep Learning** refers to advanced algorithms that identify complex patterns in data, and they can be beneficial for objectives such as predicting treatment response and detecting depression;
- 4. Neural Networks (NN)** are artificial neuron networks that are capable of modelling intricate input-output relationships and detecting data patterns.

Miner et al. (2019) neatly lay out four different approaches of AI-human integration within mental health services.

- 1. No AI Use:** On one extreme is the belief that conversational AI should not be used at all, as this could have negative unintended repercussions for both patients and clinicians.
- 2. Human-Delivered, AI-Informed:** In this method, a listening device is brought into the room and linked to software that recognizes clinically significant information, including symptoms or interventions, and communicates this information to the patient or physician.
- 3. AI-Delivered, Human-Supervised:** In this method, patients communicate with a conversational AI directly in order to make diagnoses or administer treatment. Either a human clinician would evaluate patients and assign particular duties to conversational AI, or a human clinician would monitor interactions between patients and front-line conversational AI.

4. Fully AI-Driven: On the other end is the complete removal of the human clinician out of the therapeutic relationship, with AI outperforming even the most skilled and compassionate medical professional.

AI Psychotherapy tools such as Woebot, Wysa, Replika, Tess, Youper, Ellie offer AI-based psychotherapeutic interventions anytime and anywhere. For example, Tess is a psychotherapy chatbot that evaluates patients' language, emotions, and behaviour using natural language processing, machine learning, and deep learning. It then modifies responses to provide individualised therapy. Tess has shown success in reducing anxiety and depression, especially among students, by using a range of treatment techniques, such as cognitive behavioural therapy and motivational interviewing. Although Tess is not a substitute for professional treatment, it is an auxiliary tool that increases access to mental health support and offers educational materials for the development of self-help skills.⁸

Since conversational AI is not constrained by the time or attention of human clinicians, it may be able to assist in addressing the issue of inadequate clinician availability, with the psychiatrist-to-population ratio in India being less than the recommended 3 per 100,000.¹ Long-standing barriers to accessing mental health care may be resolved if conversational AI proves to be successful and well-received by both patients and therapists. Among these are the capacity to serve rural communities with the increasing penetration of smartphones and encourage more participation from those who might find conventional talk therapy stigmatising.²

ML and NLP could be potentially beneficial tools for utilising untapped data in mental health. AI-based systems, particularly an attention-based multi-modal MRI fusion model, have been shown by Zheng et al.⁹ to be useful in the diagnosis and treatment of major depressive disorder. Their research demonstrates how AI-powered continuous monitoring can increase diagnosis precision and offer real-time solutions, ultimately improving the treatment of anxiety and depression.

Patients and providers may gain from forming a therapeutic alliance with conversational AI. Clinicians' attention and skills could be used more wisely if they let conversational AI handle time-consuming, repetitive activities. The job satisfaction of therapists may be enhanced by reducing the amount of labour that causes burnout, such as repetitive duties carried out with little autonomy. AI is thus reforming psychotherapy by increasing efficiency through task automation. Chatbots increase the accessibility and affordability of therapy. Clinicians are already using texting services to deliver mental health interventions,¹⁰ which demonstrates a willingness by patients and clinicians to test new approaches to patient-clinician interaction.

Challenges and future directions

Trust and safety might be the very first issues of concern or barriers for testing these new approaches. These issues require attention from both legal regulatory bodies and professional ethic boards. Despite the fact that national mental health programs have not specifically approved AI in psychotherapy, its potential is well known, and thorough and secure evaluations of its uses are needed.

Globally, regulatory agencies are being called upon to keep an eye on AI in healthcare, and groups such as the United Nations and the World Health Organisation are establishing guidelines for AI governance and regulation. The WHO has delineated "Key AI Principles" including preserving autonomy, promoting human well-

being and the public interest, fostering responsibility and accountability and more, as the foundational pillars to guide the application of AI in healthcare. In the US, there is no federal legislation over AI. The Trump Administration revoked the Biden Administration's Executive Order on AI (2023) upon taking office in 2025. This new executive order emphasizes on removing the barriers to innovation and posing the US as a global leader in AI. At the state level, several of them have proposed or enacted their own laws. Professional associations like the American Psychological Association (APA)¹¹ have come up with ethics updates, but these remain guidelines and are not legally or professionally binding. The European Union (EU) enforced the EU Artificial Intelligence Act in 2024, the most comprehensive framework enacted so far to govern AI use in healthcare. It offers a risk-based approach to regulate AI, classifying the systems ranging from those posing minimal risks to those with unacceptable threats. The Indian Council of Medical Research (ICMR) has established guidelines for the application of AI in biomedical research and healthcare.¹ The use of AI in the mental health sector in India also demands a comprehensive oversight.

Successful treatment requires that patients self-disclose personal information, including delicate subjects including trauma, substance abuse, sexual history, forensic background, and suicidal ideas. Professional standards and laws have been developed to set limits on what a clinician in a traditional psychotherapeutic setting can and cannot disclose (such as suicidal and homicidal tendencies that call for mandatory reporting). Simultaneously, software that supports clinical tasks has come under scrutiny for separating clinicians from patient care. This risk is particularly significant in the field of mental health because therapy frequently deals with very personal issues. Some of the time-consuming, repeated actions that clinicians perform with patients, including reviewing symptoms or obtaining their history, are really ways for them to build rapport and connect with their patients.

Technology has hazards associated with privacy, bias, coercion, liability, and data sharing that may cause patients to suffer both anticipated (such as being denied health insurance) and unexpected/unmitigated harm (such as through unauthorised data access or breaches).² Stringent addressal of such pertinent questions of data ownership, use and control is thus warranted to safeguard sensitive personal information, such as medical histories, therapy session records, and behavioural data. For instance, Talkspace and other AI-powered mental health platforms comply with the US Health Insurance Portability and Accountability Act (HIPAA), 1996 (Olawade et al., 2024).¹²

Bias in algorithms and training data further perpetuates ongoing injustices, prolonging the unfair disparities in access to healthcare and treatment outcomes. This might mean misdiagnoses, insufficient treatment, or even worsening of people's mental health conditions. It is thus essential to incorporate bias-mitigation techniques such as involving a wide range of participants (mental health professionals and marginalized communities) in the development and assessment of AI tools.¹²

AI presents difficulties with being reliable in complex therapeutic settings. It is unable to cross complex trauma, modify therapy plans with the same care and accuracy as a human therapist, or recognise small emotional shifts.^{13,14} According to Fitzpatrick et al.,¹⁵ it may miss or fail completely in mental emergencies, providing predefined responses when genuine assistance is needed. Apart from their clinical limitations, these tools also present ethical issues - individuals without digital access have been left out,^{16,17} cultural nuances are frequently

interpreted incorrectly because of training data that is Western-centric,¹⁸ and highly personal data may be shared without appropriate consent.¹⁹

The societal burden of treating mental health issues may be lessened if AI is able to diagnose or treat patients. Compared to professionals who come and go from training facilities, conversational AI may have a longer-lasting interaction with a patient. Furthermore, patients shouldn't expect a human therapist to permanently recall whole conversations due to the inherent limitations of human memory. This ability contrasts sharply with conversational AI, which can hear, recall, share, and analyse conversations for as long as it wants. Patient perceptions of AI skills may influence treatment choices and consent to data sharing because humans and machines have such disparate abilities.²⁰

Although AI may not possess the technological sophistication as of yet to tackle the drawbacks currently noticeable to us, some models such as the GPT 4.5 are increasingly getting closer to their potential of "passing" the Turing Test (able to hold seemingly human conversations). What this means for its integration with mental health care is that it needs to be approached both sensitively and scientifically—with curiosity and concern—essentially a balance between technical progress and ethical duties.

Conclusion

Rather than viewing AI as a replacement for traditional therapy, it is more realistic and productive to consider it a complementary tool—one that supports, augments, and extends the capacities of trained professionals, but does not replace them. As these technologies evolve, their integration must be guided by ethical caution, cultural awareness, clinical judgment, and an unwavering commitment to patient well-being.

Disclosures

RS, TG, SB and NKS report no relevant disclosures. The views expressed by the authors are their own and do not necessarily reflect the views of the institutions and organizations which the authors serve. All authors share first author status.

Acknowledgments

None.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

- Beg MJ, Verma M, Verma MK. Artificial intelligence for psychotherapy: A review of the current state and future directions. *Indian J Psychol Med.* 2024.
- Miner AS, Shah N, Bullock KD, et al. Key considerations for incorporating conversational AI in psychotherapy. *Front Psychiatry.* 2019;10:746.
- Boucher EM, Gray AE, Reyna L, et al. Can artificial intelligence support human flourishing? The challenges and opportunities of AI-enabled positive psychology. *J Posit Psychol.* 2021;16(4):504–510.
- Banerjee D. The COVID-19 outbreak: Crucial role the psychiatrists can play. *Asian J Psychiatry.* 2020;50:102014.
- United Nations. Policy Brief: COVID-19 and the Need for Action on Mental Health. Published 2020. Accessed July 8, 2025.
- Joshi A, Tammana S, Babre T, et al. Psychosocial response to COVID-19 pandemic in India: Helpline counsellors' experiences and perspectives. *Couns Psychother Res.* 2020;21(1):19–30.
- Rosen CS, Glassman LH, Morland LA. Telepsychotherapy during a pandemic: A traumatic stress perspective. *J Psychother Integr.* 2020;30(2):174–187.
- Bocheliuk V, Shevtsov A, Pozdniakova O, et al. Effectiveness of psycho-correctional methods and technologies in work with children who have autism. *Syst Rev J Intellect Disabil Diagn Treat.* 2023;11(1):10–20.
- Zheng G, Zheng W, Zhang Y, et al. An attention-based multi-modal MRI fusion model for major depressive disorder diagnosis. *J Neural Eng.* 2023;20(6):066005.
- Schaub MP, Wenger A, Berg O, et al. A web-based self-help intervention with and without chat counseling to reduce cannabis use in problematic cannabis users: Three-arm randomized controlled trial. *J Med Internet Res.* 2015;17(10):e232.
- American Psychological Association. Artificial Intelligence and Mental Health Care. Published 2024. Accessed July 8, 2025.
- Olawade DB, Wada OZ, Odetayo A, et al. Enhancing mental health with Artificial Intelligence: Current trends and future prospects. *J Med Surg Public Health.* 2024;3:100099.
- Vaidyam AN, Wisniewski H, Halamka JD, et al. Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *Can J Psychiatry.* 2019;64(7):456–464.
- Bendig E, Erb B, Schulze-Thuesing L, et al. The next generation: Chatbots in clinical psychology and psychotherapy to foster mental health – A scoping review. *Psychotherapie.* 2019;64(1):51–66.
- Fitzpatrick KK, Darcy A, Vierhile M. 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 Ment Health.* 2017;4(2):e19.
- Torus J, Myrick KJ, Rauseo-Ricupero N, et al. Digital mental health and COVID-19: Using technology today to accelerate the curve on access and quality tomorrow. *JMIR Ment Health.* 2020;7(3):e18848.
- Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (WYSA) for digital mental well-being: Real-world data evaluation. *JMIR Mhealth Uhealth.* 2018;6(11):e12106.
- Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? *Behav Brain Sci.* 2010;33(2-3):61–83.
- Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. *JAMA Netw Open.* 2019;2(4):e192542.
- Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: Mixed-methods systematic review. *JMIR Ment Health.* 2019;6(10):e14166.