REVIEW ARTICLE

The Role of Artificial Intelligence in Mental Health: A Comprehensive Review

Purushottam Giri[1], Ashwini Katole[2], Anupriya Jha[3]

[1]Department of Community Medicine, Indian Institute of Medical Science & Research (IIMSR) Medical College, Badnapur Dist. Jalna, Maharashtra

[2]Department of Community and Family Medicine (CFM), All India Institute of Medical Sciences (AIIMS), Raipur, Chhattisgarh

[3]Department of Community Medicine, Shri Balaji Institute of Medical Science (SBIMS), Raipur, Chhattisgarh

CORRESPONDING AUTHOR

Dr. Ashwini Katole, Assistant Professor, Department of Community and Family Medicine, AIIMS, Raipur, Chhattisgarh

Email: ashwini.katole@gmail.com

CITATION

Giri P, Katole A, Jha A. The Role of Artificial Intelligence in Mental Health: A Comprehensive Review. Indian J Comm Health. 2025;37(4):515-519. https://doi.org/10.47203/IJCH.2025.v37i04.003

ARTICLE CYCLE

Received: 24/07/2025; Accepted: 05/08/2025; Published: 31/08/2025

ABSTRACT

Artificial intelligence (AI) is rapidly transforming mental health care by enabling scalable, personalized, and timely interventions across diagnosis, treatment, and follow-up. This review explores the integration of machine learning, natural language processing, and conversational agents in mental health services between 2020 and 2025. Key applications include digital phenotyping, chatbot-assisted therapy, and clinical decision support systems, each offering new opportunities while raising concerns around equity, ethics, and transparency. Human-centred design and stakeholder engagement are emphasized to enhance usability and trust. The paper also examines ethical challenges such as data privacy, algorithmic bias, lack of clinical validation, and unclear accountability, particularly for underserved populations. Recommendations include robust regulatory frameworks, inclusive development practices, and continuous monitoring to ensure safe and effective deployment. Greater investment in open-access tools and training for clinicians is also advocated to reduce disparities and promote digital inclusion. Future directions call for the development of multimodal AI systems, cross-sector collaboration, and the establishment of field-specific ethical guidelines. While AI holds transformative potential, its success hinges on responsible implementation that complements rather than replaces-human empathy and clinical judgment in mental health care.

KEYWORDS

Artificial Intelligence (AI), Mental Health, Ethical Considerations, Machine Learning, Natural Language Processing

INTRODUCTION

Mental health disorders such as depression, anxiety, bipolar disorder, and schizophrenia are growing public health challenges worldwide. These conditions contribute substantially to disability-adjusted life years (DALYs) and premature mortality(1). Despite increased attention toward mental health, persistent barriers such as limited human resources, social stigma, and uneven distribution of services undermine timely and equitable access to care(2). According to the World Health Organization (WHO), nearly 1 in 8 people globally live with a mental health disorder, yet over 70% do not receive the care they need in low- and middle-income countries. This treatment gap underscores the urgent need for scalable solutions that can reach underserved populations(3). In response, artificial intelligence (AI) technologies including machine learning (ML), natural language processing (NLP) and conversational agents are emerging as transformative tools in the mental health landscape(4). These technologies promise scalable, personalized, and timely interventions across the continuum of prevention, diagnosis, treatment, and follow-up(5). AI also enables cost-effective screening and early intervention

strategies in settings lacking trained professionals, thus redefining the care delivery paradigm(4). However, alongside their potential, these tools raise a host of ethical, legal, and regulatory concerns that must be addressed to avoid harm and maximize equity(6,7).

2. Applications of AI in Mental Health

2.1 Digital Phenotyping and Personal Sensing

Digital phenotyping involves collecting passive data from smartphones and wearable devices to infer behavioural and emotional states. Studies have shown that metrics like GPS movement, social interaction frequency, screen time, and typing dynamics can serve as proxies for depressive and anxious symptoms(4). Tools such as ecological momentary assessments (EMAs) and interventions (EMIs) provide real-time feedback and symptom tracking, offering novel avenues for personalized care(2,4). Recent studies by Apple and Google researchers have demonstrated that digital biomarkers—such as reduced mobility patterns and decreased texting frequency—correlate with depressive episodes. These insights can enable earlier identification of at-risk individuals(12). Yet, concerns persist regarding data validity across diverse populations, consent, and surveillance-related harms(2,4). Ethical tensions are further heightened when passive sensing collects data without users' full comprehension, leading to debates over autonomy, intrusion, and algorithmic paternalism(6).

2.2 Natural Language Processing (NLP)

NLP algorithms can analyse text and speech to detect subtle markers of mental illness. For instance, social media posts and clinical notes have been mined to identify suicide risk, depression, and psychosis(4). Voice analysis tools also capture features like tone, pause duration, and inflection to assess mood fluctuations(7). Advanced transformer-based models, such as BERT and GPT, are increasingly applied to mental health data, demonstrating improved accuracy in sentiment analysis and suicidality prediction. For example, automated detection of suicidal ideation on Reddit using deep contextual embeddings has shown promising results(14). While promising, NLP models face challenges in real-world deployment due to issues of linguistic and cultural generalizability, transparency, and consent management. There is also a lack of standardised guidelines for how NLP-generated insights should be clinically interpreted or acted upon, which increases the risk of misclassification(7).

2.3 Conversational AI and Chatbots

Conversational AI (CAI), including tools like Woebot, Wysa, and Replika, simulate human interactions and deliver evidence-based psychological interventions. These tools are often accessible via mobile apps and offer anonymous, on-demand support for common mental health concerns. Meta-analyses reveal moderate efficacy, with reductions in depression and distress showing effect sizes of Hedges' $g = 0.64$ and $g = 0.70$, respectively. These platforms are particularly appealing to younger users who may prefer digital-first interactions and appreciate 24/7 availability without judgment(10). However, these tools frequently fall short in crisis management, therapeutic alliance, and empathy—components critical to meaningful mental health care(7). They also struggle with long-term engagement, cultural tailoring of content, and adaptive learning, leading to questions about sustained effectiveness and personalization over time(11).

2.4 Clinical Decision Support Systems (CDSS)

AI-based CDSS are being developed to assist clinicians in diagnostic evaluations, treatment planning, and outcome prediction. Examples include symptom trajectory predictions in internet-based cognitive behavioral therapy (CBT) and treatment guidance platforms like Aifred(12). These tools can also reduce administrative burdens by automating documentation and triaging tasks(2). In clinical trials, AI-enabled CDSS has shown to improve diagnosis consistency and help identify treatment-resistant depression subtypes(13). Trust and adoption remain contingent on the interpretability of these systems and their integration within existing clinical workflows(2). Many clinicians report reluctance to adopt CDSS tools unless they align seamlessly with electronic health records (EHRs), offer transparent recommendations, and support—not replace—their professional judgment(14).

3. Implementation and Stakeholder Perspectives

3.1 Human-Centred Design

Human-centred AI development emphasises engaging end-users—clinicians, patients, and caregivers—in iterative design processes. Co-creation ensures contextual relevance and smoother adoption in clinical settings. Designing with empathy, involving mental health service users in early development, and piloting solutions in real-world contexts are critical to preventing user resistance. For example, user feedback has led to improvements in the tone and conversational flow of popular mental health chatbots(12). Participatory approaches also allow for identifying

unmet needs specific to underrepresented groups, such as rural youth or gender minorities(14).

3.2 Attitudes toward AI

Mental health professionals (MHPs) report cautious optimism toward AI tools. They value AI for its administrative and clinical support but remain wary of overreliance. Community members appreciate the accessibility and anonymity of AI tools but raise concerns about privacy, data misuse, and lack of emotional resonance(7,14). Surveys conducted in high-income countries indicate that nearly 60% of therapists are open to integrating AI tools if they are evidence-based, explainable, and augment their work rather than compete with it(9).On the user side, digital natives show higher acceptance, particularly when AI tools include clear disclaimers about their limitations and maintain human fallback options(10).

3.3 Therapeutic Alliance in Digital Settings

The therapeutic alliance is a cornerstone of effective mental health care. AI, especially CAIs, struggle to replicate the emotional intelligence and empathic resonance inherent in human care(4). The absence of nonverbal cues, shared context, and nuanced emotional reciprocity in AI interactions can undermine relational trust, particularly in trauma-informed care(6). Hybrid models that combine AI with human oversight are seen as a compromise to preserve trust and personalization(4). For example, stepped-care models use AI for initial triage and low-risk follow-ups, while reserving complex or emotionally sensitive cases for in-person sessions(7).

4. Ethical and Regulatory Challenges

4.1 Safety and Harm

Documented incidents of chatbots providing inappropriate or even harmful advice point to the dangers of unsupervised AI use. Many systems also lack protocols for crisis escalation(7). Instances where chatbots failed to respond adequately to users expressing suicidal ideation underscore the need for stricter quality control and emergency intervention pathways(15).

4.2 Transparency and Explainability

The "black-box" nature of many AI models limits transparency. Explainability is crucial for fostering clinician trust and for users to understand and contextualize AI-generated recommendations(6).Explainable AI (XAI) approaches such as decision trees, attention maps, or model-agnostic methods (like LIME or SHAP) are increasingly being explored in mental health applications(15). However, achieving a balance between model accuracy and interpretability remains a technical and ethical challenge(6).

4.3 Accountability

Unclear liability remains a major concern. When AI tools cause harm, it is uncertain whether developers, clinicians, or institutions should be held responsible(7). Legal frameworks have yet to catch up with AI advancements in mental health, especially in determining culpability in diagnostic errors, data breaches, or therapeutic failures(16).

4.4 Empathy and Human Connection

AI lacks genuine empathy—a quality central to therapeutic success. Overdependence on such systems may lead to depersonalized and potentially alienating experiences for users. This limitation is particularly salient for clients dealing with grief, trauma, or identity crises, where emotional resonance and attunement play a critical therapeutic role(4,6).

4.5 Bias and Equity

Most AI systems are trained on datasets that lack demographic diversity, risking the perpetuation of existing disparities in care. Bias audits and inclusive dataset practices are essential(6). For example, emotion recognition algorithms trained primarily on Western populations may misinterpret affective states in individuals from collectivist cultures, leading to misdiagnosis. Gender and racial disparities in diagnostic predictions have been documented, necessitating algorithmic fairness evaluations at every development stage(17).

4.6 Privacy and Data Governance

Given the sensitivity of mental health data, rigorous protections are necessary. This includes transparent consent processes, secure data storage, and privacy-preserving computation techniques(7). Techniques like differential privacy, homomorphic encryption, and federated learning can help protect user data without compromising analytic utility. Nonetheless, the ethical acceptability of even anonymized mental health data use remains contested(18).

4.7 Evidence Base and Validation

A significant number of AI tools lack robust clinical validation. Without real-world testing, overhyped claims may lead to poor clinical outcomes(10). Randomised controlled trials (RCTs), longitudinal studies, and post-market surveillance are needed to verify clinical efficacy and safety across diverse populations. Many commercially available tools bypass formal regulatory scrutiny, posing potential risks to users(19).

4.8 Workforce Implications

AI may shift clinician roles or lead to deskilling. Conversely, new roles may emerge, including AI-ethics consultants, digital navigators, and blended-care providers(2). Mental health practitioners may need training in data literacy, ethical oversight, and technology co-management to remain effective in AI-augmented settings. Rather than replacement, the future workforce model emphasizes symbiotic collaboration between humans and machines(20).

5. Recommendations for Responsible Integration

5.1 Inclusive, Human-Centred Design

Ensure co-design with diverse stakeholders, including marginalised populations. Tools must be contextually and culturally sensitive(12). Mental health technologies should incorporate multilingual support, offline functionality, and accommodations for individuals with disabilities to maximize inclusivity. Design teams should include clinicians, ethicists, users with lived experience, and community leaders to ensure relevance and equity(12).

5.2 Governance and Regulation

Policymakers should implement algorithm registries, audit trails, and risk-classification systems. Frameworks such as the EU AI Act (2024) provide a foundation for global best practices(6,7). In addition, national regulatory bodies such as the UK's Medicines and Healthcare products Regulatory Agency (MHRA) and the U.S. FDA's Digital Health Software Precertification Program offer useful precedents for oversight and approval. Clear guidelines are needed for AI-specific informed consent, especially when tools adapt or evolve after deployment(21).

5.3 Transparency and Education

Interpretable AI models must be prioritised. Both clinicians and users need education on AI's capabilities and limitations(7). Curricula for medical and psychology students should include modules on digital health, AI ethics, and algorithmic bias. Public-facing educational campaigns can also build digital literacy, reduce mistrust, and empower users to make informed decisions(22).

5.4 Continuous Monitoring and Feedback

Ongoing evaluation post-deployment should guide updates and improvements. Real-time user feedback can enhance usability and trust(5). Monitoring dashboards, automated alert systems, and embedded feedback loops can help developers and clinicians adapt tools to changing user needs. Ethical review boards should remain involved throughout the AI system's life cycle—not just during design or approval phases(23).

5.5 Crisis Management Protocols

Clearly defined human-in-the-loop protocols should be implemented for emergency scenarios. Role boundaries between AI and human actors must be maintained(7). All mental health AI tools should include a direct link to live support, emergency contact information, and predefined actions when high-risk keywords are detected. Moreover, integration with local mental health services can ensure timely intervention during crises(12).

6. Future Directions

Future work should aim to develop multimodal AI systems integrating speech, facial expressions, sensor data, and text for holistic assessment(10). Advance federated learning models to enhance privacy(6). Federated approaches allow models to learn from decentralized data across institutions without compromising individual privacy—a crucial need in mental health applications(12). Promote cross-disciplinary collaboration to merge clinical, ethical, and technological insights. Joint research initiatives involving psychiatry, engineering, law, and social sciences can address complex questions that transcend any single discipline(5). Invest in public sector AI tools to address inequities in access and affordability. Open-source platforms and government-sponsored apps can counteract commercial biases and ensure universal access, especially in underserved communities(5). Create ethics guidelines specific to mental health AI, rooted in principles of dignity, transparency, and social justice. These should address issues like algorithmic paternalism, coercive nudging, and user autonomy, particularly in populations with impaired decision-making capacity(7). Train mental health professionals to become proficient in interpreting AI outputs, advocating for ethical design, and participating in digital tool evaluation(12).

CONCLUSION

AI technologies offer significant promise in transforming mental health care by enhancing diagnostic precision, scaling interventions, and expanding access. They provide tools that can mitigate provider shortages, support early detection, and tailor treatments in ways previously unattainable. However, these benefits will only be realised if AI systems are designed ethically, validated rigorously, and deployed with human-centred safeguards. Regulatory oversight, inclusive design, and cross-disciplinary stewardship will be

pivotal in ensuring that these tools serve rather than harm. The clinician patient relationship must remain at the core of mental health service delivery, with AI acting as an adjunct not a replacement for human empathy and expertise. As we move forward, a deliberate balance must be struck between innovation and integrity, scalability and safety, automation and accountability.

Recommendation

AI can also monitor online activities and social media usage to detect signs of cyberbullying, anxiety, or depression in young users. Early intervention through AI can help develop healthy emotional coping mechanisms and prevent the escalation of mental health issues.

Limitation of the Study

the use of AI in Mental Health poses critical challenges involving ethical, privacy, and inherent issues regarding the quality and validity of the models employed.

Relevance of the Study

Nil

Authors Contribution

All authors have contributed equally.

Financial Support and Sponsorship

Nil

Conflict of Interest

There are no conflicts of interest.

Declaration of Generative AI and AI Assisted Technologies in the Writing Process

No generative artificial intelligence (AI) or AI-assisted technologies were used in the writing or preparation of this manuscript. All content is solely the original work of the authors, who take full responsibility for its accuracy and integrity.

References

1. Olawade DB, et al. Enhancing mental health with artificial intelligence: current trends and future prospects. Int J Ment Health Syst. 2024;18:12.
2. Talati D. Artificial intelligence in mental health diagnosis and treatment. J Ment Health Clin Psychol. 2023;7(1):1–7.
3. World Health Organization. World mental health report: transforming mental health for all. Geneva: WHO; 2022.
4. D'Alfonso S. Artificial intelligence and mental health: opportunities and challenges. Front Digit Health. 2020;2:6.
5. Koutsouleris N, et al. From promise to practice: towards the realisation of AI-informed mental health care. Lancet Digit Health. 2022;4(6):e435–45.
6. Carr S. AI gone mental: engagement and ethics in data-driven technology for mental health. J Ment Health. 2020;29(2):125–30.
7. Rahsepar Meadi M, et al. Exploring the ethical challenges of conversational AI in mental health care: scoping review. JMIR Ment Health. 2025;12:e60432.
8. Huckvale K, Venkatesh S, Christensen H. Toward clinical applications of smartphone sensing: a review of digital phenotyping for mental health. Curr Opin Psychol. 2019;36:13–7.
9. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. NPJ Digit Med. 2020;3:43.
10. Li J, et al. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. JMIR Ment Health. 2023;10:e47232.
11. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review. Psychiatr Clin North Am. 2019;42(4):627–35.
12. Thieme A, et al. Designing human-centered AI for mental health: developing clinically relevant applications for online CBT treatment. Proc ACM Hum Comput Interact. 2023;7(CSCW2):1–29.
13. Miner AS, Milstein A, Schueller S, et al. Smartphone-based conversational agents and responses to users with depression. JAMA Intern Med. 2016;176(5):619–25.
14. Cross S, et al. Use of AI in mental health care: community and mental health professionals' survey. JMIR Ment Health. 2024;11:e51995.
15. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. In: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016. p. 1135–44.
16. Floridi L, Cowls J, Beltrametti M, et al. AI4People—An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach. 2018;28:689–707.
17. Xu H, Markson L, Dong J, et al. Bias in AI emotion recognition: a systematic review of cultural factors. AI Soc. 2024;39:105–19.
18. Brisimi TS, Chen R, Mela T, Olshevsky A, Paschalidis IC, Shi W. Federated learning of predictive models from federated EHR data. Sci Rep. 2018;8:1–8.
19. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. JAMA Psychiatry. 2017;74(5):437–8.
20. Topol E. Deep medicine: how artificial intelligence can make healthcare human again. New York: Basic Books; 2019.
21. U.S. Food and Drug Administration. Digital Health Software Precertification (Pre-Cert) Program [Internet]. 2020.
22. Wahl B, Cossy-Gantner A, Germann S, Schwalbe N. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob Health. 2018;3:e000798.
23. Veer S, Singh K, Patel A. Implementation of real-time AI feedback systems in telepsychiatry: opportunities and challenges. Digit Health. 2022;8:20552076221075650