Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

ML Powered Guardian System for Mental Health Risk Prediction

[1]Prajakta Kamble,[2]Dr.D.S.Waghole

[1] Department of Computer Engineering,
[1]Jayawantrao Sawant College of Engineering, Pune, India

Abstract - Artificial intelligence (AI)-enabled digital interventions are increasingly used to expand access to mental health care. This PRISMA-ScR scoping review maps how AI technologies support mental health care across five phases: pre-treatment (screening), treatment (therapeutic support), post-treatment (monitoring), clinical education, and population-level prevention. Methods: We synthesized findings from 36 empirical studies published through January 2024 that implemented AI-driven digital tools, including large language models (LLMs), machine learning (ML) models, and conversational agents. Use cases include referral triage, remote patient monitoring, empathic communication enhancement, and AI-assisted psychotherapy delivered via chatbots and voice agents. Results: Across the 36 included studies, the most common AI modalities included chatbots, natural language processing tools, machine learning and deep learning models, and large language model-based agents. These technologies were predominantly used for support, monitoring, and self-management purposes rather than as standalone treatments. Reported benefits included reduced wait times, increased engagement, and improved symptom tracking. However, recurring challenges such as algorithmic bias, data privacy risks, and workflow integration barriers highlight the need for ethical design and human oversight. Conclusion: By introducing a four-pillar framework, this review offers a comprehensive overview of current applications and future directions in AI-augmented mental health care. It aims to guide researchers, clinicians, and policymakers in developing safe, effective, and equitable digital mental health interventions.

Keywords: **Digital mental health; artificial intelligence; conversational agents; large language models; machine learning; chatbots; AI-assisted psychotherapy; mental health screening; remote patient monitoring; scoping review**

1. INTRODUCTION

In recent decades, mental health disorders have surged despite economic and technological progress, with barriers such as stigma, cost, and professional shortages continuing to hinder treatment access [1–3]. Digital health technologies offer promise in addressing these challenges [4], with teletherapy, mental health apps, and computerized cognitive behavioral therapy showing effectiveness [5]. The COVID-19 pandemic further accelerated the use of digital mental health tools, revealing unmet needs at the intersection of technology and psychotherapy [6,7]. This growing demand has sparked increased interest in the role of conversational artificial intelligence in mental health applications [8,9], with research exploring its potential to enhance psychotherapy and improve care delivery [10–13]. Researchers have increasingly focused on leveraging conversational artificial intelligence to support psychotherapy effectiveness, as the mental health sector grapples with rising demand and the need for innovative solutions [12–16]. Since the public release of ChatGPT (Version GPT-4, United States of America) in early 2023, it has become the first conversational artificial intelligence tool to achieve global mainstream use, reshaping approaches to learning, communication, and problem solving [8,17]. Research on ChatGPT has expanded rapidly across disciplines, especially in education, medicine, and psychology, highlighting its growing potential to advance mental health services [10–13].

Artificial intelligence-driven digital interventions in mental health refer to software systems or mobile applications that embed artificial intelligence techniques to deliver, support, or evaluate mental health services [6,12].These include conversational artificial intelligence agents that interact with users through natural language, ranging from simple FAQ style or rule-based chatbots to more advanced multi-turn dialogue systems capable of handling complex communication tasks [18,19]. Natural language processing techniques enable these agents to parse user input, detect sentiment, and extract key emotional cues [20]. Machine learning models, such as classification and regression algorithms, and deep learning networks, such as convolutional and recurrent neural networks, power predictive and monitoring tools to classify diagnoses, forecast risks, and tailor treatment recommendations based on user data [20]. Large language models such as GPT and BERT, which belong to a subclass of deep learning models built on transformer architectures with self-attention mechanisms, expand capabilities by generating and comprehending coherent and context-rich text, opening new possibilities for nuanced therapeutic dialogue and personalized content creation [18,19]. Artificial intelligence has emerged as a promising tool to augment human therapists [12], although its adoption challenges traditional care models and raises concerns regarding efficacy, ethics, privacy, and the interpretation of

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

human mental health experiences [21]. This review aims to provide a comprehensive analysis of conversational artificial intelligence in mental health care by mapping empirical evidence across different clinical phases. It offers insights into current applications, challenges, and future opportunities.

2. METHODS

2.1. Research Aims

Various types of AI technologies are utilized within the broad context of mental health care. Many of these technologies are specifically linked to the conversational AI interface, which engages users to provide a wide range of support. Various technologies, including AI chatbots, different language models, prediction modeling, sentiment analysis, and recommender systems, have been implemented into health care settings [22–25]. These technologies are making significant advancements in mental health care by improving diagnostic accuracy, enhancing personalized treatment, providing insights and recommendations to clinicians, tailoring services to individual needs, and offering accessible and cost-effective mental health support to everyone [26–29]. As the field of computer science continues to progress, there is a transformative opportunity for the mental health care field to understand and apply these technologies to their services effectively. However, it is crucial to approach this integration thoughtfully, maintaining standards of care and prioritizing patient-centric approaches. There is a need for a scoping review examining how different AI technologies are being used in mental health, their impacts, ethical considerations, and practical aspects of combining AI with human care across various settings.

This review aims to present a framework for the existing state of integrating AI into mental health services in a manner that maximizes benefits and minimizes risks. The summarization can serve as an overview for those interested in researching and developing solutions for these issues. To realize their full promise, integration must be guided by evidence on efficacy, ethical safeguards, and alignment with clinical workflows. Accordingly, this scoping review maps the current landscape of AI-driven digital interventions in mental health, proposes a four-pillar mapping to organize empirical findings, and identifies practical barriers and enablers to maximize benefits and minimize risks. Our primary goals are to chart which AI modalities are deployed in each care phase, summarize their proven outcomes and reported limitations, and outline strategic directions for future research and policy.

2.2 Design and Scope of the Study

This scoping review adhered to the PRISMA-ScR guidelines [30] to ensure a rigorous, transparent, and reproducible methodology for mapping the use of AI-driven digital interventions in mental health care. We define our scope as all conversational AI agents (from rule-based/FAQ chatbots to ML-powered multi-turn systems and transformer-based LLMs) and related predictive/monitoring models (NLP and ML/DL algorithms) deployed across five phases: (1) pre-treatment (screening and triage), (2) treatment (therapeutic support), (3) post-treatment (follow-up and monitoring), (4) clinical education, and (5) general improvement and prevention. Conducting this review, we examined how each technology is applied, its demonstrated outcomes (e.g., accuracy, engagement, and health gains), and its limitations, thereby offering a unified life cycle framework for AI in mental health.

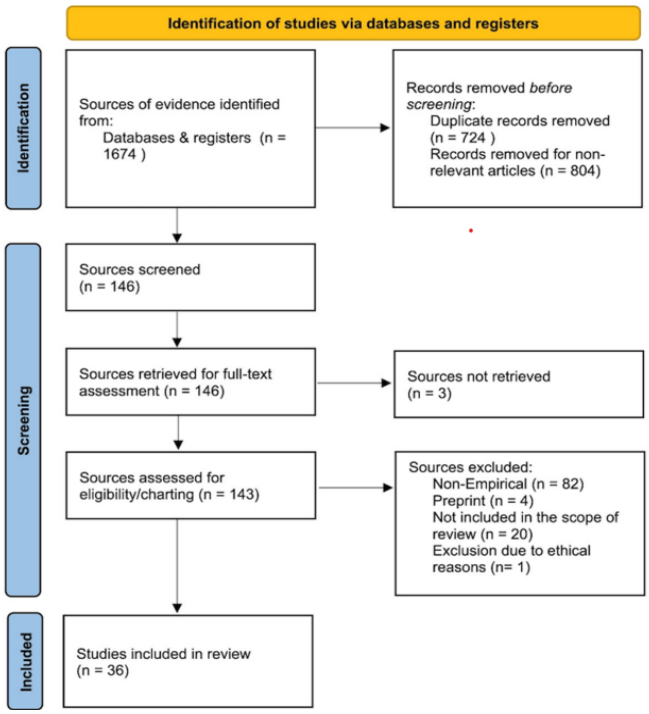

Figure 2.1. PRISMA-ScR flow chart

2.3 Search Strategy and Data Extraction

A customized data-charting form was developed to capture key information: study phase (application scenarios), AI technology, setting, outcomes, and limitations from each included article. It captured key details such as the objectives, AI technologies utilized, main findings, and conclusions. This structured approach allowed for a thorough and organized analysis of the empirical evidence on conversational AI in mental health care. Two reviewers independently extracted all study-level data; any discrepancies were discussed and resolved by consensus adjudication prior to synthesis.

3. RESULT

3.1 Summary of Reviewed Results

The review process included 36 articles that showcased survey results of users, clinical students, and mental health professionals' perceptions toward AI applications. Additionally, all 36 articles evaluated the efficacy of specific aspects of AI applications. To organize the reviewed studies, we classified them based on the primary AI technologies employed. Given the considerable overlap between categories, we focused on highlighting the main types of technologies without reporting precise counts. The technologies identified across the studies include AI chatbots, conversational agents, natural language processing (NLP) tools,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

large language models (LLMs), machine learning (ML) models, deep learning (DL) models, and AI-based prediction systems. We acknowledge that many studies employed hybrid approaches or combinations of these technologies. To improve clarity, we report general trends and key examples without quantifying the exact number of studies per category.

3.2. Applying Natural Language Processing

Through the applications developed by natural language processing (NLP), machine learning (ML), and deep learning (DL), AI tools empower clinical services and treatments and make mental health services much more accessible [25,35]. First, NLP, with mainly applications of chatbots and AI agents, is extensively employed to enhance the interactivity and efficacy of using conversational agents to help deliver mental health care [25,37,40]. Many research-oriented AI-driven chatbots or agents, such as “Hailey”, “MYLO”, and “Limbic Access”, are utilizing sophisticated NLP algorithms to deeply understand human language and initiate, respond, and engage in meaningful conversations related to users’ well-being. Conversations empowered by NLP techniques, such as emotion detection and sentiment analysis, can effectively help provide mental health support and offer computerized therapies [29,36]. The capabilities of NLP in engaging users through text and voice interactions foster mental health interventions with no limits on time and space. Many AI-based chatbots leveraged NLP to provide mental health support through conversations [23,40]. An AIenhanced cognitive behavioral therapy (CBT) chatbot named Hailey used NLP to analyze user text inputs and, specifically, detect emotions and train users to give empathic responses for facilitating peer-to-peer communications, and an agent called TEO provided similar techniques to enhance stress reduction [34,48].

Moreover, preliminary research on ChatGPT highlighted LLM-based conversational AI’s potential to expand access to mental health services to all populations due to its incredible capabilities in understanding human language. It surpasses capabilities in identifying emotions, tailoring interventions for conditions such as borderline personality disorder (BPD) and sensory processing disorder (SPD), and offering advice that aligns with primary health care guidelines for depression. In addition to the opportunities NLP brought, some NLP models in psychiatry still demonstrated significant biases related to religion, race, gender, nationality, sexuality, and age, which highlighted the need for further enhancing the technology [56].

3.3. Applying Machine Learning

Machine learning (ML) has proven to be a powerful tool in predicting mental health outcomes and enhancing diagnostic accuracy, ultimately aiming to improve treatment efficiency and recovery outcomes. By leveraging predictive analytics and classification models, ML provides clinicians with valuable decision-support systems and enables the creation of personalized treatment plans for clients [24,57]. For example, machine learning models, including logistic regression, ridge regression, and LASSO regression, have been used to develop AI-based assessment tools that can accurately predict mental disorders based on responses to a mental health assessment tool like the SCL-90-R [32]. This tool can diagnose mental disorders with an impressive accuracy of 89% using just 28 questions. Regarding remote patient monitoring, ML algorithms analyze data from devices monitoring vital signs and physical activity to identify subtle changes indicative of worsening mental health conditions, such as depression or anxiety [24,36]. This capability is crucial for timely intervention in these conditions, where even minor changes can be serious indicators of patient distress.

3.5. Applying Deep Learning

Deep learning (DL) algorithms, a subset of machine learning, learn complex patterns directly from data, enabling accurate predictions and analyses for innovative applications such as real-time emotional state monitoring and predictive analytics for treatment outcomes. DL contributes to improving online cognitive behavioral therapy (CBT) and art psychotherapy by customizing mental health treatments to cater to individual needs [31,43]. In the realm of iCBT, DL algorithms and recurrent neural networks (RNNs) are employed to analyze anonymous patient data [31]. They detect patterns that accurately forecast treatment outcomes, assisting in identifying mental health issues and the customization of therapy for more effective and personalized interventions [31,35]. Similarly, in art psychotherapy, DL models with co-attention mechanisms are revolutionizing the evaluation of art therapy [43]. These models assess stress and mood levels based on multiple data points. DL’s capacity to interpret complex emotional expressions and provide insights that align closely with therapeutic goals [42,43]. Through these advancements, DL serves as a technological tool and acts as a bridge to compassionate, precise, and personalized mental health care. Having mapped the core AI modalities above, we now turn to how these tools (chatbots, predictive models, and LLMs),are deployed at each phase of mental health care.

Key Findings in the Applications in Mental Health Care Here, we answer RQ2 by showing how those same AI modalities are deployed across the five clinical phases (pre-treatment, treatment, post-treatment monitoring, clinical education, and prevention), highlighting which technologies are most prevalent in each phase and where critical gaps remain.

The review identified various applications of AI in mental health care that cater to diverse users, ranging from patients and clinicians to the general public and psychology students. These applications serve different purposes, including providing computerized therapies to patients, offering early-stage mental health support, assisting clinicians with diagnosis and treatment, and enhancing learning for psychology students. Overall, the studies showcased the positive impact of AI technology in improving mental health care and emphasized the significant potential of these applications to revolutionize the industry. Interestingly, most of the studies focused on how these AI-powered applications can complement and enhance the existing services provided by clinicians rather than replacing them. Figure 2 illustrates the four pillars of AI applications in mental health care, demonstrating their

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

extensive utilization across various stages of support and treatment. The four pillars encompass four key stages: pre-treatment, treatment, post-treatment, and general improvement and prevention. Artificial intelligence is leveraged throughout these stages to enhance various aspects of mental health services, creating a continuous source for optimized care and support. In the pre-treatment phase, AI expedites assessment, facilitates initial diagnosis, and aids in referral. During treatment, AI refines diagnoses, personalizes treatment plans, predicts outcomes, and delivers AI-based therapeutic interventions. Posttreatment involves leveraging AI for remote monitoring and risk evaluation. The general improvement and prevention stage focuses on providing proactive mental health support to the broader population through AI-enabled resources.

Throughout the four pillars framework, AI technologies benefit multiple stakeholders (see Figure 3). Patients gain access to fast-track services and more effective, personalized interventions. Clinicians can make enhanced, data-driven decisions. Health organizations improve efficiency and treatment efficacy. The general public benefits from increased access to low-cost, AI-powered mental health resources. This integrated model showcases the vast potential of AI in revolutionizing mental health care. In the following sections, we turn to the existing literature to explore examples of AI implementation within each stage of the model.

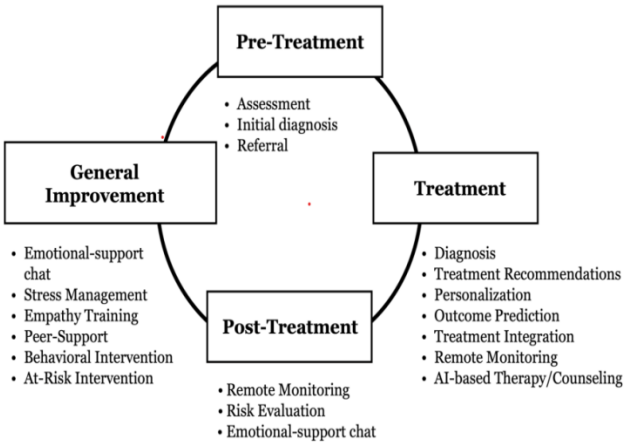

Figure 3.5 Four pillars framework where AI-driven interventions are deployed

CONCLUSION

This scoping review synthesized findings from 36 studies on artificial intelligencedriven digital interventions across screening, treatment, monitoring, clinical education, and prevention in mental health care. Our mapping of chatbots, natural language processing, machine learning, deep learning, and large language models highlighted growing evidence of artificial intelligence's contributions to expanding access, enhancing symptom monitoring, and supporting personalized interventions. At the same time, challenges such as algorithmic bias, data privacy risks, and integration barriers underscore the need for ethical design, transparent model development, and human oversight. While previous review studies have provided valuable insights into specific aspects of AI in mental health, such as diagnostic applications of ChatGPT, ethical and regulatory discussions, or evaluations of chatbots for anxiety treatment [73–80], these have often focused on single technologies, clinical phases, or conditions. In contrast, the current scoping review offers a broader synthesis by mapping multiple AI modalities across five clinical phases. By linking conceptual insights with empirical outcomes, this review complements and extends prior work, providing an integrated reference for research, practice, and policy

Future research should prioritize multi-site evaluations, longitudinal studies in diverse populations, and rigorous assessments of safety, privacy, and equity impacts. Collaboration among clinicians, artificial intelligence developers, policymakers, and patients will be essential to ensure artificial intelligence systems are clinically effective, ethically sound, and socially equitable. By offering an empirical map of artificial intelligence applications across the mental health care continuum, this review provides a foundation for guiding research, practice, and policy toward responsible integration of artificial intelligence in mental health services.

REFERENCES

[1] 1. Coombs, N.C.; Meriwether, W.E.; Caringi, J.; Newcomer, S.R. Barriers to healthcare access among U.S. adults with mental health challenges: A populationbased study. SSM Popul. Health 2021, 15, 100847.
[2] Hidaka, B.H. Depression as a disease of modernity: Explanations for increasing prevalence. J. Affect. Disord. 2012, 140, 205–214.
[3] Uutela, A. Economic crisis and mental health. Curr. Opin. Psychiatry 2010, 23, 127–130.
[4] Torous, J.; Myrick, K.J.; Rauseo-Ricupero, N.; Firth, J. Digital mental health and COVID-19: Using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment. Health 2020, 7, e18848.
[5] Arean, P.A. Here to stay: Digital mental health in a post-pandemic world— Looking at the past, present, and future of teletherapy and telepsychiatry. Technol. Mind Behav. 2021, 2, e00073.
[6] Friis-Healy, E.A.; Nagy, G.A.; Kollins, S.H. It is time to REACT: Opportunities for digital mental-health apps to reduce mentalhealth disparities in racially and ethnically minoritized groups. JMIR Ment. Health 2021, 8, e25456.
[7] Prescott, M.R.; Sagui-Henson, S.J.; Chamberlain, C.E.W.; Sweet, C.C.; Altman, M. Real-world effectiveness of digital mental-health services during the COVID-19 pandemic. PLoS ONE 2022, 17, e0272162.
[8] Wu, T.; He, S.; Liu, J.; Sun, S.; Liu, K.; Han, Q.L.; Tang, Y. A brief overview of ChatGPT: History, status-quo and potential future development. IEEE/CAA J. Autom. Sin. 2023, 10, 1122–1136.
[9] Lattie, E.G.; Stiles-Shields, C.; Graham, A.K. An overview of and recommendations for more accessible digital mental-health services. Nat. Rev. Psychol. 2022, 1, 87–100.
[10] Adeshola, I.; Adepoju, A.P. The opportunities and challenges of ChatGPT in education. Interact. Learn. Environ. 2023, 32, 6159–6172.
[11] Biswas, S.S. Role of ChatGPT in public health. Ann. Biomed. Eng. 2023, 51, 868–869.
[12] D'Alfonso, S. AI in mental health. Curr. Opin. Psychol. 2020, 36, 112–117.
[13] Su, S.; Wang, Y.; Jiang, W.; Zhao, W.; Gao, R.; Wu, Y.; Tao, J.; Su, Y.; Zhang, J.; Li, K.; et al. Efficacy of Artificial Intelligence-assisted psychotherapy in patients with anxiety disorders: A prospective, national multicentre randomized controlled trial protocol. Front. Psychiatry 2022, 12, 799917.
[14] Sedlakova, J.; Trachsel, M. Conversational artificial intelligence in psychotherapy: A new therapeutic tool or agent? Am. J. Bioeth. 2022, 23, 4–13.
[15] Henson, P.; Wisniewski, H.; Hollis, C.; Keshavan, M.; Torous, J. Digital

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

mentalhealth apps and the therapeutic alliance: Initial review. BJPsych Open 2019, 5, e15.
[16] Vilaza, G.N.; McCashin, D. Is the automation of digital mental health ethical? Front. Digit. Health 2021, 3, 689736.
[17] Roumeliotis, K.I.; Tselikas, N.D. ChatGPT and Open-AI models: A preliminary review. Future Internet 2023, 15, 192.
[18] Adamopoulou, E.; Moussiades, L. An overview of chatbot technology. In Artificial Intelligence Applications and Innovations: AIAI 2019; Springer: Cham, Switzerland, 2020; pp. 261–280.
[19] Bayani, A.; Ayotte, A.; Nikiema, J.N. Transformer-based tool for automated fact-checking of online health information: Development study. JMIR Infodemiol. 2025, 5, e56831.
[20] Hang, C.N.; Yu, P.D.; Chen, S.; Tan, C.W.; Chen, G. MEGA: Machine-learningenhanced graph analytics for infodemic-risk management. IEEE J. Biomed. Health Inform. 2023, 27, 6100–6111.