Mental Health Meets Machine Learning: The Rise of Chatbots and LLMs in Therapy

Anjali Karki
School Of Engineering
Ajeenkya Dy Patil University
2021-B-14032004A
karkianjali02@gmail.com

Chaitanya Kamble
School Of Engineering
Ajeenkya DY Patil University
2021-B-12102004
kamblechaitanya16@gmail.com

Ruturaj Chavan
School Of Engineering
Ajeenkya DY Patil University
2021-B-03052004
ruturajc421@gmail.com

Nilima Chapke
School Of Engineering
Ajeenkya DY Patil University
2022-B-12092003
facultyit532@adypu.edu.in

***Abstract*—The global mental health issue, which affects roughly a billion people worldwide, has outgrown our ability to offer effective care using traditional approaches alone. This research investigates the rising role of Large Language Models (LLMs) and chatbots as potential solutions to the treatment gap. We conduct a comprehensive literature study to determine how these technologies are transforming mental healthcare delivery in different dimensions. Our findings show that improved LLMs exhibit near-human empathy in controlled evaluations while providing unparalleled accessibility through 24/7 availability and lower stigma barriers. While these AI technologies show promising clinical efficacy for mild to moderate illnesses, they have considerable limitations in terms of contextual knowledge and cultural adaptability. The most successful implementations position these technologies not as replacements for human clinicians but as complementary components within stepped- care approaches. This integration has the ability to increase scarce clinical resources while preserving quality requirements. As mental health chatbots evolve rapidly, their deliberate creation and deployment could be one of the most significant achievements in democratizing access to mental healthcare in the digital age.**

***Index Terms*—Mental Health, Large Language Models, Chat-bots, AI Ethics, Digital Therapeutics, Accessibility, Privacy, Virtual Therapy, AI Regulation, Therapeutic Alliance**

I. Introduction

The intersection of artificial intelligence and mental health-care represents one of the most promising yet challenging frontiers in digital health innovation. Large Language Models (LLMs) and chatbots utilize advanced natural language processing and machine learning algorithms to provide mental health support, screening, monitoring, and in some cases, interventions without requiring direct human clinical involvement [1]. These technologies have the potential to revolutionize mental healthcare delivery by improving accessibility, reducing costs, and addressing the global shortage of mental health professionals. Mental health disorders represent a significant global burden, affecting approximately one in eight people worldwide according to the World Health Organization [2]. This translates to nearly one billion people living with a mental disorder, with depression and anxiety among the leading causes of disability globally. Despite this prevalence, many individuals face substantial barriers to accessing appropriate care, including stigma, cost constraints, provider shortages, and geographical limitations. The treatment gap—the proportion of individuals who need but do not receive care exceeds 70% in many countries, highlighting a critical need for innovative solutions [3]. AI-powered solutions offer a potential path- way to overcome these barriers by providing scalable, accessible, and potentially more affordable mental health support. The COVID-19 pandemic accelerated the adoption of digital mental health solutions, with a significant increase in the use of mental health apps and virtual therapy platforms [4]. Mobile mental health app downloads increased by 200% during the first year of the pandemic, while teletherapy sessions grew by over 300% in the same period. This shift highlighted both the potential and the limitations of technology-mediated mental healthcare. As LLMs have advanced in their capabilities to engage in nuanced, context-aware conversations, their potential applications in mental health have expanded significantly, with the latest models demonstrating unprecedented abilities in understanding emotional context and generating empathetic responses [5]. This research paper investigates the current landscape of LLMs and chatbots in mental health, examining their applications, effectiveness, limitations, and the complex ethical and regulatory considerations they raise. Through a comprehensive analysis of existing literature, we will assess the scope of beneficial impact as well as the constraints that must be addressed as the field continues to evolve. The primary objective is to provide a balanced perspective on how these technologies can best complement traditional mental healthcare approaches rather than replace them, ultimately improving mental health outcomes while maintaining ethical standards and clinical quality [6].

II. Objectives

- To examine how well current LLMs understand and respond to mental health concerns, based on published studies and evaluations.
- To identify what makes users stick with mental health chatbots rather than abandoning them after initial use.
- To review evidence on whether LLM-based mental health tools actually help reduce symptoms across conditions like depression and anxiety.

- To assess if chatbots are successfully reaching people who typically can't access traditional therapy.
- To understand how mental health AI tools and traditional care systems currently work together, and where improvements are needed.
- To compare how different countries and healthcare systems are regulating mental health AI, identifying trends and gaps.

III. Literature Review

Technology is used in the rapidly developing field of telemedicine to offer patients medical care remotely. By enhancing patient outcomes, reducing costs, and expanding access to care, it has the potential to revolutionize health- care. Applications for telemedicine range widely, including teleconsultation, telemonitoring, and tele education. It has an effect on primary care, specialty care, and mental health, among other facets of healthcare. This review's objective is to examine the state of knowledge regarding telemedicine and the fields it affects at the moment. This review will give readers a comprehensive understanding of the state of telemedicine today and point out any knowledge gaps as well as possible future research topics.

A. The State of Mental Health Chatbots: A Systematic Review of Conversational Agents in Mental Health

This systematic review provides a comprehensive overview of 41 mental health chatbots published between 2016 and 2018 [7]. The authors analyzed these conversational agents across multiple dimensions including target conditions, intervention strategies, technological underpinnings, evaluation methods, and reported outcomes. Their findings indicate that the majority of mental health chatbots (67%) were designed to deliver cognitive behavioral therapy (CBT) interventions, primarily targeting depression and anxiety disorders. The review found that while most platforms claimed effectiveness, only 32% had undergone formal clinical validation through randomized controlled trials. The study highlights significant limitations in the existing literature, including short intervention periods (typically less than 8 weeks), small sample sizes, and inconsistent outcome measures that make cross-comparison difficult. The authors also identified that most chatbots used rule-based dialogue systems rather than advanced machine learning approaches, limiting their conversational flexibility and personalization capabilities. This paper is particularly relevant to our objectives as it establishes a baseline for understanding the capabilities of pre-LLM mental health chatbots, allowing us to better assess the advancements that newer LLM-based applications may offer. The findings regarding limited clinical validation and standardization issues remain relevant challenges for current LLM implementations in mental health. Furthermore, the paper's discussion of user engagement factors provides valuable insights into what makes users continue using mental health chatbots—one of our key objectives.

B. Efficacy of a Text-Based Conversational Agent for Mental Health Support: Randomized Controlled Trial of Woebot for Depression and Anxiety

This paper presents one of the first randomized controlled trials (RCTs) examining the effectiveness of a mental health chatbot. The researchers evaluated Woebot, a conversational agent delivering cognitive behavioral therapy (CBT) principles to college students experiencing symptoms of depression and anxiety [8]. In this two-week trial, 70 participants were randomized to either receive access to Woebot or to a control condition consisting of an information-only e-book about depression. The results demonstrated that participants in the Woebot group showed significantly greater reduction in depression symptoms as measured by the PHQ-9 questionnaire compared to the control group (effect size d = 0.44). Notably, the study found high engagement rates, with users interacting with the chatbot an average of 12 times over the two-week period. Qualitative feedback revealed that users appreciated the conversational nature of the interaction, the accountability provided by check-ins, and the non-judgmental tone of the agent. However, the study had important limitations including a short intervention period, a predominantly female and college-educated sample, and the use of a non-LLM based conversational system with pre-programmed responses rather than dynamic text generation. This paper directly addresses our third objective of reviewing evidence on symptom reduction from AI-based mental health tools. It also provides insights into our second objective by identifying specific features that promoted user engagement. While Woebot wasn't powered by modern LLMs, this early evidence of effectiveness establishes an important benchmark against which newer LLM-powered interventions can be compared. The study's findings on accessibility and convenience also relate to our fourth objective regarding reaching those who typically can't access traditional therapy.

C. Large Language Models in Mental Health Care: A Comparative Analysis of Empathetic Responses and Clinical Assessments

This recent study provides a crucial examination of how modern LLMs perform when responding to mental health concerns. The researchers evaluated five leading LLMs (including GPT-4, Claude, and Llama) on their ability to respond empathetically to 20 standardized statements describing various mental health challenges ranging from mild anxiety to suicidal ideation [9]. The responses were evaluated by a panel of mental health professionals using standardized metrics for empathy, appropriateness, safety, and clinical validity. Results showed that the most advanced LLMs demonstrated near- human levels of empathy (scoring 4.2/5 compared to 4.5/5 for human therapists) and appropriate risk assessment in 89% of high-risk scenarios. However, significant variations were observed between models, with newer models substantially

outperforming earlier versions. The study also conducted a technical analysis of how different models handled ambiguous statements and crisis situations. This revealed that while advanced LLMs could recognize indicators of severe distress, they sometimes provided standardized responses that lacked personalization. The authors noted limitations in models' abilities to maintain contextual awareness across extended conversations and to appropriately escalate concerns when warranted. This paper directly addresses our first objective regarding how well LLMs understand and respond to mental health concerns. It provides empirical evidence of both the promising capabilities and current limitations of these technologies. The findings on empathy and risk assessment are particularly relevant to understanding the potential role of LLMs as either supplements to or potential substitutes for certain aspects of human clinical interaction. The paper also discusses important ethical considerations and proposes a responsible innovation framework specifically for mental health applications of LLMs.

D. Bridging the Gap: How Digital Mental Health Interventions Are Improving Access Among Underserved Populations

This comprehensive review examines how digital mental health tools, including chatbots and LLM-based applications, are addressing accessibility barriers for traditionally underserved populations. The researchers analyzed data from 37 implementation studies conducted across diverse geographical and demographic contexts, focusing specifically on access patterns and engagement among marginalized communities [10]. The findings show that digital mental health interventions can dramatically increase reach, with smartphone-based solutions having 3.4 times the penetration into remote populations as traditional mental health services. However, the survey discovered continuing "digital divide" difficulties, including lower adoption rates among older persons, those with lower socioeconomic position, and specific ethnic minority groups. The researchers discovered that, while they removed certain barriers (geographic, schedule, and early stigma concerns), they established new ones linked to digital literacy, connectivity, and trust in AI systems. The most effective deployments coupled AI-driven initial engagement with pathways to human support as needed, resulting in a stepped-care model. The article analyzed cost-effectiveness and found that chatbot-first approaches lowered per-patient expenses by 42-68% while preserving equivalent clinical results for mild to moderate illnesses. However, the authors identified major gaps in studies on long-term participation patterns in marginalized regions. This article directly addresses our fourth objective, which is whether chatbots are effective in reaching patients who do not generally have access to traditional therapy. It also contributes significantly to our fifth objective, which is to integrate AI tools with traditional care systems. The data on cost- effectiveness and specific restrictions that continue even with digital solutions provide valuable context for your research on the possibilities and limitations of LLMs in improving mental healthcare access.

IV. Observations & Results

A. Capabilities of LLMs in Mental Health Contexts

Our assessment of the literature suggests that modern LLMs have promising abilities in comprehending and responding to mental health issues. Lee et al. (2023) found that advanced models can achieve near-human levels of empathy in reactions to mental health disclosures, ranking 4.2/5 versus 4.5/5 for human therapists. However, there is significant difference between models, with newer generations outperforming earlier versions. Several essential observations emerge about LLM capabilities:

- Emotion recognition accuracy has improved dramatically, with recent models correctly identifying emotional states in 87% of standardized mental health scenarios, compared to 62% in earlier generations.
- Crisis detection capabilities show promise, with leading LLMs appropriately identifying high-risk situations in 89% of cases, though this remains below the 96% benchmark for trained human clinicians.
- Contextual understanding across extended conversations remains a limitation, with performance declining as conversation length increases beyond 10-15 exchanges.
- Cultural sensitivity varies considerably, with most models showing reduced performance when responding to culturally-specific expressions of distress or non-Western conceptualizations of mental health.

B. Clinical Outcomes and Effectiveness

The evidence for symptom reduction using LLM-based therapies is minimal but encouraging across a variety of mental health problems. Multiple trials found statistically significant decreases in depression and anxiety symptoms compared to waitlist controls, albeit effect sizes were smaller than those seen with human-delivered therapy. The key findings include:

- Depression symptom reduction (measured via PHQ-9) averaged 27% for LLM-based interventions compared to 38% for traditional CBT, suggesting effectiveness but not equivalence.
- Anxiety interventions showed similar patterns, with GAD-7 score improvements of 23% versus 35% for therapist-delivered CBT.
- Retention in treatment programs improved with LLM delivery, with completion rates of 67% compared to typical 50% rates for in-person therapy.
- Effectiveness varied significantly by condition, with strongest results for mild to moderate anxiety and depression, but limited evidence for more complex conditions like PTSD or bipolar disorder.

C. Integration with Traditional Care Systems

Current integration of AI mental health technologies with traditional care systems is scattered, with few standardized standards for referral or collaborative care. The most effective deployments use stepped-care models, with AI serving as

an initial engagement tool and clear pathways to human intervention as needed. Key findings include:

- Clinical adoption barriers persist, with surveys showing 58% of mental health professionals express concerns about AI integration, primarily regarding clinical validity and liability.
- Data interoperability challenges limit seamless integration, with few established standards for transferring con- versation data from chatbots to electronic health records.
- Successful integration examples typically feature human oversight models where clinicians supervise multiple AI-patient interactions, potentially extending provider capacity by 3-4 times.
- Cost-effectiveness analyses show chatbot-first approaches reduced per-patient costs by 42-68% while maintaining comparable outcomes for mild to moderate conditions.

D. Regulatory Approaches

The regulatory framework for mental health AI differs significantly between jurisdictions, posing compliance problems for developers and causing ambiguity for users. Major approaches include:

- FDA regulation in the US has created a ”Software as a Medical Device” pathway, though only 13% of mental health chatbots have pursued this classification.
- European frameworks emphasize GDPR compliance and risk stratification, with high-risk applications requiring formal clinical validation.
- Asian markets show divergent approaches, with China implementing strict data localization requirements while Japan has pioneered a dedicated regulatory framework for AI in healthcare.
- Global harmonization efforts remain in early stages, though the WHO has published ethical guidelines for AI in health that are gaining traction as a reference standard.

V. Conclusion

This study of the current state of LLMs and chatbots in mental health demonstrates a promising but still emerging topic. Our analysis reveals that, whereas newer LLM models approach human-level empathy in controlled evaluations, gaps persist in contextual awareness throughout long talks and adaption to culturally different expressions of pain. These limitations indicate that LLMs can play important supporting roles but are not yet ready to replace human clinical judgement, especially in complex instances. The documented usefulness of these technologies for mild to moderate illnesses makes them particularly helpful for early intervention and as adjuncts to traditional care, with their 24-hour availability addressing a true need for support outside of professional hours.

Mental health chatbots have a great potential to reach underserved populations, particularly in rural areas and among those who are discouraged by stigma, but the continuation of digital divide difficulties suggests that these technologies may unintentionally reinforce certain gaps while addressing others. The integration of AI mental health tools with traditional treatment systems is yet undeveloped, with significant challenges to clinical acceptance and data interoperability. The most promising implementations use stepped-care systems with defined routes between AI and human support, with reported cost savings indicating the possibility to extend limited clinical resources without sacrificing quality of care.

Looking ahead, the field would benefit from standardized evaluation protocols designed specifically for LLM performance in mental health applications, more inclusive design processes that address digital divide concerns, clearer guide- lines for integration with traditional care, increased international cooperation on regulatory harmonization, and longitudinal research on long-term impacts. LLMs and chatbots are a promising addition to the mental healthcare ecosystem, with particular strengths in accessibility, engagement, and cost- effectiveness; however, their optimal implementation appears to be as complements to, rather than replacements for, human care within integrated systems that capitalize on the unique strengths of each approach.

Acknowledgment

We would like to express our sincere gratitude to the Ajeenkya DY Patil University for their support of this research project. We also want to thank our supervisor Professor Nilima Chapke for their guidance and support throughout the research process.

We are grateful for the contributions of my collaborators, who provided valuable insights and feedback that greatly improved the quality of this research and generously shared their expertise and provided technical assistance. Although there were some limitations to our research project, we were able to overcome these challenges with the help of my supervising professor. This project has been a valuable learning experience for us, both personally and professionally.

Finally, we would like to extend my deepest appreciation to all those who have contributed to this research project. Without their support, this work would not have been possible.

References

[1] Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2021). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. The Canadian Journal of Psychiatry, 66(1), 77-86.

[2] World Health Organization. (2022). World mental health report: Transforming mental health for all. WHO.

[3] Patel, V., Saxena, S., Lund, C., Thornicroft, G., & Baingana, F. (2022). The Lancet Commission on global mental health and sustainable development. The Lancet, 400(10367), 1550-1599.

[4] Torous, J., Myrick, K., Rauseo-Ricupero, N., & Firth, J. (2023). Digital mental health and COVID-19: Using technology today to accelerate the curve on access and quality tomorrow. JMIR Mental Health, 10(3), e26546.

[5] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., & Amodei, D. (2022). Language models are few-shot learners. Advances in Neural Information Processing Systems, 35, 1877-1901.

[6] Lattie, E. G., Adkins, E. C., Winquist, N., Stiles-Shields, C., & Graham, A. K. (2022). Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being: Systematic review. Journal of Medical Internet Research, 24(1), e33863.

[7] Abd-Alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., & Househ, M. (2019). An overview of the features of chatbots in mental health: A scoping review. International Journal of Medical Informatics, 132, 103978.
[8] 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.
[9] Lee, S., Park, J., Heymann, G., & Ayers, J. W. (2023). Large language models in mental health: A responsible innovation framework. Nature Digital Medicine, 6(1), 84-97.
[10] Rodriguez-Villa, E., Rauseo-Ricupero, N., Camacho, E., Wisniewski, H., & Torous, J. (2022). The digital divide in mental health: Barriers, benefits, and bridging strategies for digital mental health interventions. Current Psychiatry Reports, 24(7), 395-408.