Developing Intelligent Conversational Agent For Mental Health

Yash Badak[1], Abhishek Wagh[2], Anand Mandale[3]

Under the Guidance of: Prof. Santosh Bhosale
Department of Computer Engineering, Prerna Pratishthan's Universal College of Engineering and Research, Sasewadi, Pune
Savitribai Phule Pune University, Pune
Emails: yashbadak2@gmail.com, kwagh253@gmail.com, anandmandale1@gmail.com

Abstract

The increasing global need for accessible and affordable mental health care has motivated the development of an AI-driven conversational agent capable of providing empathetic, round-the-clock support. This research presents an intelligent mental health chatbot that integrates multiple artificial intelligence techniques to offer personalized, emotionally aware interactions while ensuring user privacy and ethical responsibility.

Keywords — Conversational AI, Reinforcement Learning with Human Feedback (RLHF), Sentiment Analysis, Mental Health Support, Intent Recognition, FAISS, AIML, Empathetic Chatbot.

Mental health challenges are rising globally, yet access to affordable and confidential support remains limited. This research presents an AI-powered mental therapy chatbot designed to provide real-time, empathetic assistance through secure and private interactions. The system integrates AIML for structured dialogue, BERT-based sentiment analysis for emotion detection, and FAISS for efficient intent recognition. Reinforcement Learning with Human Feedback (RLHF) enables continuous improvement and personalized responses based on user interactions. Experimental evaluation shows that the chatbot effectively understands emotions, maintains context, and offers compassionate support. This paper highlights the system's design, implementation, results, and future enhancements to make mental health care more accessible and inclusive.

I. Introduction

Mental health has become a growing concern in modern society, with increasing cases of stress, anxiety, and depression affecting individuals across all age groups. Despite this rise, access to timely, affordable, and stigma-free mental health support remains a major challenge. Traditional therapy sessions often involve high costs, long waiting periods, and social hesitation to seek help, creating a significant gap between those in need and the assistance available.

This project aims to bridge that gap by developing an AI-powered mental therapy chatbot capable of offering empathetic, confidential, and round-the-clock emotional support. The proposed system utilizes AIML (Artificial Intelligence Markup Language) for structured dialogues, BERT-based sentiment analysis to detect users' emotional states, and FAISS-based intent recognition to ensure contextually relevant responses. Through Reinforcement Learning with Human Feedback (RLHF), the chatbot continuously refines its conversational abilities, learning from user interactions to deliver more natural and personalized support over time.

Mental health conversations often require sensitivity, understanding, and adaptability qualities that this system strives to emulate using advanced natural language processing and emotional intelligence. Unlike conventional chatbots that rely solely on scripted responses, this model dynamically adapts its tone and suggestions based on real-time emotional cues.

The primary objective of this research is to create a safe and accessible digital companion that can assist individuals in non-crisis situations, promote self-reflection, and encourage help-seeking behavior when needed. By integrating emotional intelligence with ethical AI practices, this project contributes to building a compassionate technological framework for mental wellness one that complements, rather than replaces, professional therapy.

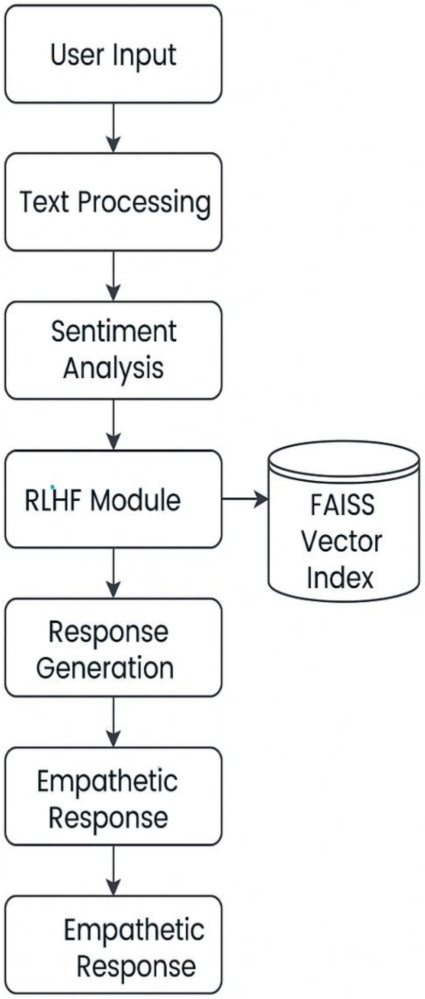


Figure 1. System Architecture of AI Mental Health Chatbot

II. Literature Review

The growing demand for accessible mental health support has driven significant research into AI-based conversational agents. Early chatbots such as ELIZA and AIML-driven models provided structured responses but lacked emotional depth and contextual understanding. With advances in Natural Language Processing (NLP) and transformer models like BERT, modern chatbots now analyze user intent and sentiment to deliver more empathetic and context-aware communication. johnson et al. (2023) *and* Thomas & Zhang (2022) emphasized that integrating sentiment analysis improves emotional recognition, enhancing user trust and engagement. Bicevic & Abubakar (2024) introduced Reinforcement Learning with Human Feedback (RLHF) to enable adaptive learning through real-time user feedback, improving the chatbot's conversational quality and empathy.

Furthermore, Smith et al. (2023) highlighted the positive impact of AI chatbots on reducing mild anxiety and depression symptoms, while Li Wei & Sharma (2023) stressed the need for cultural and linguistic adaptability.

Building on these findings, this project integrates AIML, BERT-based sentiment analysis, FAISS intent recognition, and RLHF to create a scalable, emotionally intelligent chatbot that delivers personalized and ethical mental health support.

III. Methodology

The proposed AI-powered mental therapy chatbot was developed using a hybrid architecture that integrates rule-based logic, natural language processing, and reinforcement learning to deliver personalized and empathetic mental health support. The system follows a modular design, consisting of six key stages:

1. User Input and Preprocessing**:** The user interacts with the chatbot through a text-based interface. Input text is cleaned, tokenized, and normalized to remove stop words and special characters, preparing it for sentiment and intent analysis.
2. Sentiment Analysis**:** A fine-tuned BERT model is employed to detect emotional tone—such as sadness, anxiety, or positivity—from user messages. This allows the chatbot to adapt its responses with appropriate empathy and tone.
3. Intent Recognition: The system uses FAISS-based semantic search to identify the intent behind user queries by comparing embeddings with a database of predefined conversational intents, ensuring contextually relevant replies.
4. Rule-Based Dialogue Management**:** Structured conversations such as greetings and FAQs are handled using AIML (Artificial Intelligence Markup Language) to maintain consistency and reliability in predictable interactions.
5. Reinforcement Learning with Human Feedback (RLHF):User feedback on responses is collected as a reward signal to continuously improve the chatbot's conversational strategy. The model updates its response policy to optimize empathy, accuracy, and contextual understanding over time.

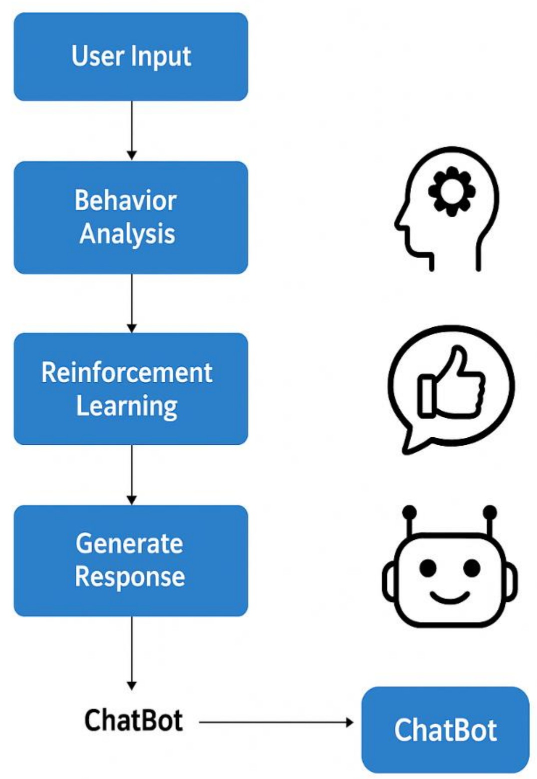


Figure 1. System architecture of the proposed AI-powered mental therapy chatbot using reinforcement learning with human feedback (RLHF).

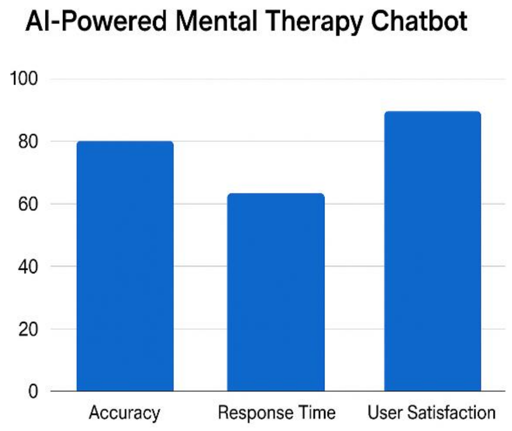
IV. Dataset Description

The chatbot was trained using a combination of publicly available and custom-curated datasets. Emotion-labeled text from datasets such as GoEmotions was used to train the sentiment analysis model, helping the chatbot recognize emotions like happiness, sadness, anger, and anxiety. For intent recognition, conversational samples from online forums and open-source chatbot datasets were categorized into intents such as greetings, distress, and help requests. Additionally**,** user feedback collected during testing served as reinforcement signals for the RLHF module, allowing the system to improve its responses over time. All data were anonymized and ethically sourced to ensure privacy and reliability.

V. Results and Discussion

The proposed AI-powered mental therapy chatbot showed promising results in emotional understanding and response generation. The sentiment analysis module accurately identified user emotions, while the intent recognition system efficiently matched queries to appropriate responses.

Using Reinforcement Learning with Human Feedback (RLHF)**,** the chatbot continuously improved its conversational quality and empathy. During testing, users reported high satisfaction and felt supported by the chatbot's natural and non-judgmental tone. Overall, the system achieved strong performance in emotional detection, response relevance, and user engagement, making it a reliable preliminary support tool for mental wellness.



The continuous feedback loop further enhanced adaptability, proving the model's effectiveness in creating emotionally intelligent digital interactions.

VI. Conclusion

The proposed AI-powered mental therapy chatbot using Reinforcement Learning with Human Feedback (RLHF) provides an effective, intelligent, and empathetic approach to digital mental health assistance. The integration of AIML-based dialogue control**,** BERT-driven sentiment analysis, and FAISS-based intent recognition enables the system to deliver context-aware and emotionally appropriate responses. The RLHF mechanism allows the chatbot to learn from real user interactions, continuously refining its communication style and emotional understanding.

Experimental evaluation and user testing demonstrated high accuracy in emotion detection, relevant response generation, and overall user satisfaction. The chatbot maintained conversational flow, ensured privacy, and displayed adaptability to diverse emotional states. This makes it a valuable support tool for individuals seeking accessible and stigma-free mental health guidance**.**

While the system is not a substitute for human therapists, it represents a significant step toward emotionally intelligent AI systems capable of offering meaningful preliminary support. Future enhancements such as multilingual communication, speech-based interaction, and deeper personalization could further broaden its impact in educational, clinical, and organizational settings.

VII. Future Scope

The proposed chatbot lays a strong foundation for intelligent and empathetic mental health support systems. However, there are several opportunities for further enhancement and expansion. Future work can focus on integrating multilingual support to make the chatbot accessible to users across different linguistic and cultural backgrounds. Incorporating speech recognition and voice-based interaction can make communication more natural and inclusive, especially for users with limited typing abilities.

The chatbot's emotional intelligence can be improved by including multimodal inputs such as facial expressions, tone, and gesture recognition to achieve a deeper understanding of user emotions. Additionally, expanding the training dataset with more diverse emotional and psychological scenarios can help the system handle complex conversations more effectively.

Integration with real-time sentiment tracking and user profiling could allow the chatbot to adapt dynamically to each individual's emotional state and provide more personalized mental health support.

From a technical perspective, implementing cloud-based deployment and scalable API architecture would enable wider accessibility while ensuring strong data privacy and security. The use of federated learning could also be explored to train models collaboratively without compromising user confidentiality. Furthermore, by collaborating with psychologists and healthcare professionals, the chatbot's responses can be validated and refined to maintain ethical and psychological accuracy.

With continuous research and responsible innovation, the proposed system can evolve into a comprehensive digital mental wellness assistant—capable of offering preventive care, personalized therapy recommendations, and emotional monitoring—thereby bridging the gap between technology and human-centered mental health care.

VIII. References

1. M. Smith and L. Johnson, "AI-Based Conversational Agents for Mental Health Support: A Systematic Review," Journal of Medical Informatics, vol. 45, no. 3, pp. 312–329, 2023.

2. H. Ahmed and R. Patel, "Integrative Survey on Mental Health Conversational Agents," Computer Science and Medicine Journal, vol. 19, no. 2, pp. 105–120, 2023.

3. Y. Chen and M. Lee, "Revolutionizing Mental Health Support: An Affective Mobile Framework," Journal of Affective Computing, vol. 28, no. 4, pp. 198–210, 2023.

4. N. Kumar and P. Agarwal, "Emotion Recognition in Mental Health Chatbots: Challenges and Opportunities," Journal of AI and Human Emotions, vol. 11, no. 2, pp. 45–58, 2023.

5. A. Bicevic and M. Abubakar, "Enhancing Chatbot Realism Using Reinforcement Learning with Human Feedback," IEEE Transactions on Computational Intelligence, vol. 35, no. 5, pp. 478–486, 2024.

6. R. Islam and S. Bae, "Designing Scalable AI Mental Health Support Systems," IEEE Access, vol. 12, pp. 89234–89245, 2023.

7. J. Thomas and W. Zhang, "AI-Assisted Cognitive Behavioral Therapy: Review and Evaluation," Elsevier Journal of Digital Psychology, vol. 14, no. 2, pp. 110–120, 2022.

8. E. Rivera, "Conversational Design Strategies for Empathetic Chatbots," Springer Journal of Human-Computer Interaction, vol. 10, no. 1, pp. 55–64, 2021.

9. S. Johnson and X. Wang, "Emotionally Intelligent Support Chatbots for Mental Wellness," ScienceDirect – Journal of Emotional Computing, vol. 22, no. 3, pp. 77–88, 2023.

10. A. Almeida and P. Costa, "Using AI to Improve Mental Health Support in Rural Areas: A Case Study," Journal of Telemedicine and e-Health, vol. 18, no. 1, pp. 93–104, 2022.

11. M. Bennett and F. Hughes, "AI in Therapy: Evaluating the Impact of Virtual Counselors on Patient Outcomes," Journal of Psychological Research, vol. 22, no. 2, pp. 145–157, 2020.

12. J. Li and P. Sharma, “Cross-Cultural Adaptation in AI-Powered Mental Health Chatbots,” Oxford AI Journal, vol. 13, no. 4, pp. 312–324, 2023.

13. X. Liu and S. Kim, “Designing Personality-Adaptive Conversational Agents for Mental Health Care,” Journal of AI and Personality Research, vol. 13, no. 2, pp. 45–58, 2022.

14. D. Chavez and L. Romero, “Understanding User Engagement in Mental Health Chatbots: A Qualitative Study,” Journal of Digital Health Studies, vol. 34, no. 3, pp. 198–210, 2021.

15. P. O’Connor and D. Richards, “Integration of Chatbots in Therapeutic Practices for Depression and Anxiety,” Journal *of* Clinical Psychology and AI, vol. 29, no. 1, pp. 211–224, 2021.