iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.81614

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

A Privacy-Aware Conversational System for Mental Health Support

Gautam Kumar[1], Ramnaresh Sharma[2]
[1]*Student*, [2]*Assistant Professor, Centre for AI, MITS Gwalior, India*

Abstract: ***Mental health issues such as stress, anxiety, and emotional isolation have become increasingly prevalent in recent years, particularly among students and young professionals. Despite the growing need for support, access to professional mental health services remains limited due to factors such as social stigma, cost, and lack of availability. This paper presents a conversational AI system designed to provide immediate and accessible emotional support through natural language interaction. The proposed system integrates Natural Language Processing (NLP) techniques with a conversational language model to interpret user input and generate context-aware responses. It incorporates an emotion detection mechanism to identify the user's emotional state and adapts its responses accordingly. A dedicated safety layer ensures that all responses remain supportive, non-harmful, and ethically appropriate. The system is implemented using Python and lightweight frameworks, making it scalable and easy to deploy. Experimental observations show that the chatbot performs effectively in handling common emotional scenarios and provides meaningful responses in real time. While the system is not intended to replace professional mental health services, it demonstrates the potential of AI-driven tools as an initial layer of emotional support.***

Keywords: Mental Health, Chatbot, NLP, Conversational AI, Emotion Detection, Human-Computer Interaction

I. INTRODUCTION

Mental health is a critical component of overall well-being, yet it is often overlooked due to societal and personal barriers. In today's fast-paced environment, individuals frequently encounter stress related to academics, careers, and personal life. While awareness regarding mental health has improved, many people still hesitate to seek help due to stigma or lack of accessible resources.

Technological advancements have opened new possibilities for addressing these challenges. Among these, conversational AI systems have gained significant attention due to their ability to simulate human-like interaction. These systems can provide a safe and private environment where users can express their thoughts without fear of judgment.

The Project is developed as an AI-based conversational assistant that focuses on emotional support. Unlike traditional chatbots that rely on predefined responses, this system uses NLP and a conversational model to generate responses dynamically based on user input. The system is designed to understand the emotional context of a message and respond in a supportive manner.

The main objective of this work is to design and implement a system that combines conversational intelligence with ethical safeguards. The system prioritizes user privacy and avoids generating harmful or misleading content. By doing so, it aims to provide a reliable and accessible support tool for individuals experiencing emotional distress.

II. LITERATURE REVIEW

The development of conversational agents has progressed significantly over the past few decades. Early systems such as ELIZA, developed in the 1960s, relied on pattern-matching techniques to simulate human conversation. Although these systems demonstrated the potential of human-computer interaction, they lacked contextual understanding and adaptability, making them unsuitable for complex or emotionally sensitive applications.

With the advancement of Natural Language Processing (NLP), modern chatbot systems have become more sophisticated. Machine learning-based approaches allow chatbots to learn from data and generate responses dynamically rather than relying solely on predefined rules. Recent developments in deep learning, particularly the use of transformer-based architectures, have significantly improved the quality of conversational systems by enabling better context understanding and response generation.

©IJRASET: All Rights are Reserved

In the domain of mental health, several AI-based applications have been developed to provide emotional support. These systems typically utilize sentiment analysis to detect user emotions and apply structured techniques to guide conversations. Some approaches incorporate principles from Cognitive Behavioral Therapy (CBT) to help users manage negative thoughts and behaviors. Studies have shown that such systems can help reduce feelings of stress and loneliness by providing immediate and accessible interaction. Despite these advancements, existing systems face several limitations. One of the major challenges is the inability of AI to fully understand complex human emotions and psychological states. Additionally, concerns related to data privacy and ethical usage remain significant, as these systems often deal with sensitive user information. Another limitation is the risk of generating inappropriate or misleading responses, especially when using generative models without proper safeguards. This Project addresses these challenges by integrating a safety-focused design with conversational intelligence. Unlike purely generative systems, it incorporates a filtering mechanism to ensure that responses remain supportive and ethically appropriate. Furthermore, the system emphasizes minimal data handling to protect user privacy. By combining emotional awareness with controlled response generation, the proposed system aims to provide a more reliable and user-centric solution for mental health support.

III. PROPOSED SYSTEM

The AI based Therapist is designed using a modular architecture that separates the system into multiple functional components. This design approach improves scalability, maintainability, and flexibility.

A. System Overview

The system consists of the following major components:

1) User Interface: Provides a platform for interaction between the user and the chatbot.
2) Text Processing Module: Prepares user input for analysis through cleaning and tokenization.
3) Emotion Detection Module: Identifies the emotional tone of the input.
4) Response Generation Module: Produces context-aware replies using a language model.
5) Safety Filtering Module: Ensures that responses remain appropriate and supportive.

B. System Workflow

The system workflow begins when a user enters a message through the interface. The input text is processed and analyzed to determine its structure and meaning. The emotion detection module evaluates the emotional tone, which helps guide the response generation process. Once a response is generated, it is passed through a safety filter that removes or modifies any potentially harmful content. The final response is then delivered to the user.

C. Design Considerations

The system is designed with the following considerations:

1) Privacy of user data
2) Ethical response generation
3) Real-time interaction capability
4) Ease of scalability

D. Technical Architecture

The technical architecture of the proposed system follows a modular and layered design, ensuring scalability, flexibility, and efficient processing of user interactions. The system begins with a user interface layer, implemented using a lightweight framework, which facilitates real-time communication between the user and the chatbot. User input is passed to the backend processing layer, where Natural Language Processing (NLP) techniques are applied for text cleaning, tokenization, and semantic understanding. This processed input is then forwarded to the emotion detection module, which analyzes the sentiment of the text to classify the user's emotional state. The architecture ensures smooth data flow between components, enabling quick and context-aware response generation. The core intelligence of the system lies in the response generation layer, which utilizes a language model to produce meaningful and empathetic replies. To maintain reliability and ethical standards, a safety filtering mechanism is integrated into the architecture, which reviews generated responses and removes any potentially harmful or inappropriate content before delivering them to the user.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Additionally, the system is designed with minimal data retention to ensure user privacy and confidentiality. The modular structure allows for easy integration of advanced models, multilingual capabilities, and future enhancements, making the architecture robust and adaptable for real-world deployment in mental health support applications.

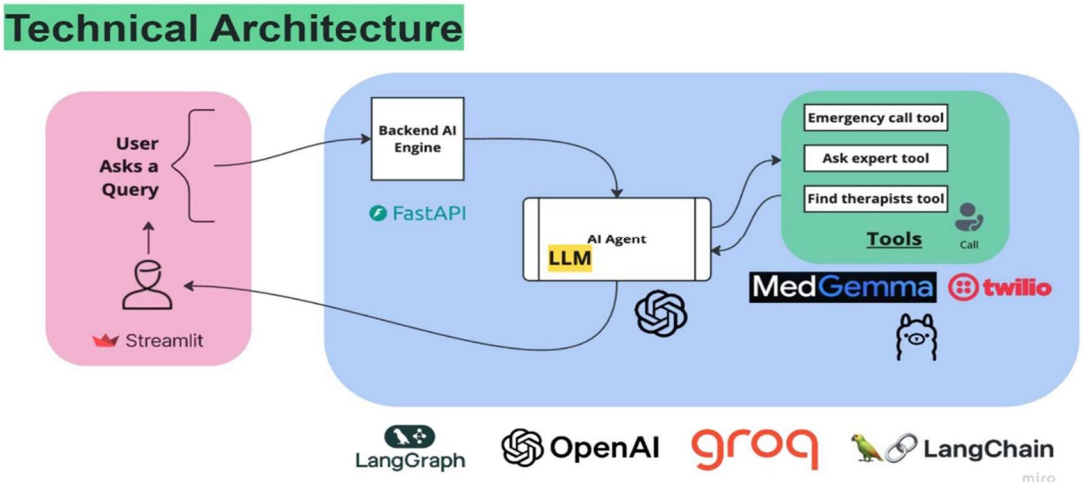


Fig.3.D Technical Architecture

IV. METHODOLOGY

The methodology involves combining NLP techniques with a conversational model to achieve effective interaction.

1) Text Preprocessing: Text preprocessing is the first step in the pipeline. It involves cleaning the input text by removing unnecessary characters, converting it to a standard format, and breaking it into tokens. This step ensures that the data is suitable for analysis.

2) Emotion Detection: Emotion detection is performed using sentiment analysis techniques. The system classifies input into categories such as positive, negative, or neutral. This classification helps the system understand the user's emotional state and respond appropriately.

3) Response Generation: The response generation module uses a language model to produce replies based on the input and detected emotion. The model ensures that the responses are coherent and contextually relevant.

4) Safety Mechanism: The safety mechanism plays a crucial role in maintaining ethical interaction. It filters responses to prevent harmful or misleading content. This is particularly important in mental health applications, where incorrect responses can have serious consequences.

V. IMPLEMENTATION

The system is implemented using Python due to its versatility and availability of powerful libraries.

1) Tools and Technologies

- Python for backend development
- Streamlit for user interface
- NLP libraries for text processing
- Language models for response generation

2) System Integration

Each module is implemented separately and then integrated into a unified system. The modular design ensures that components can be updated independently without affecting the overall system.

3) User Interface

The interface is designed to be simple and user-friendly. It allows users to interact with the chatbot in a conversational format, making the experience intuitive and accessible.

VI. RESULTS AND DISCUSSION

A. Testing Scenarios

The system was evaluated using a variety of test inputs representing different emotional states. The chatbot demonstrated the ability to generate relevant and supportive responses in most cases.

B. Performance Analysis

The response time was efficient, enabling real-time interaction. The system performed particularly well for short and clearly defined inputs. Users were able to engage with the chatbot and receive meaningful responses.

C. User Experience Evaluation

User experience evaluation shows that most users rated the system as *good* or *excellent*, indicating positive interaction. Users appreciated the simple interface, quick responses, and supportive replies. A few rated it as average due to limitations with complex inputs. Overall, the chatbot provides a reliable and user-friendly emotional support experience.

D. Limitations of the System

The system may struggle with highly complex or ambiguous inputs. Additionally, maintaining long-term context in extended conversations remains a challenge.

Despite these limitations, the results indicate that the system is effective as an initial support tool and provides a valuable platform for emotional expression.

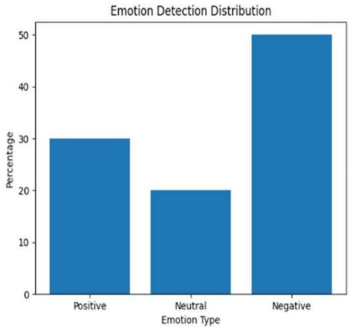


Fig.6.A

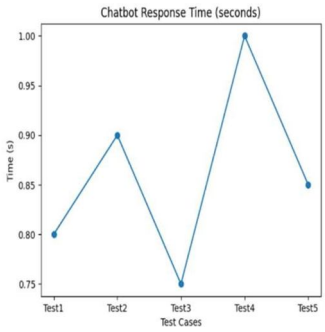


Fig.6.B

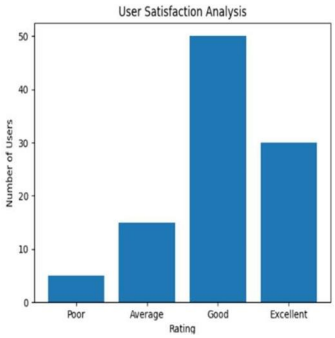


Fig.6.C

VII. APPLICATIONS AND FUTURE WORK

A. Real-World Applications

The AI based therapist has several potential applications. It can be used as a support tool for students, a self-reflection platform for individuals, and a component of larger mental health systems.

B. Future Enhancements

Future improvements may include:

- Integration of advanced models for better context understanding
- Support for multiple languages
- Voice-based interaction
- Integration with professional mental health services

These enhancements can significantly improve the effectiveness and reach of the system.

VIII. CONCLUSION

This paper presented the design and implementation of AI based therapist, an AI-based conversational system for mental health support. The system demonstrates how NLP and conversational models can be used to create a supportive interaction environment. While it is not intended to replace professional care, it provides an accessible and immediate solution for individuals seeking emotional support. The project highlights the importance of ethical considerations in AI systems, particularly in sensitive domains such as mental health.

IX. ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my project supervisor, Mr.Ramnaresh Sharma, for their valuable guidance, continuous support, and constructive suggestions throughout the development of this project. Their insights and encouragement

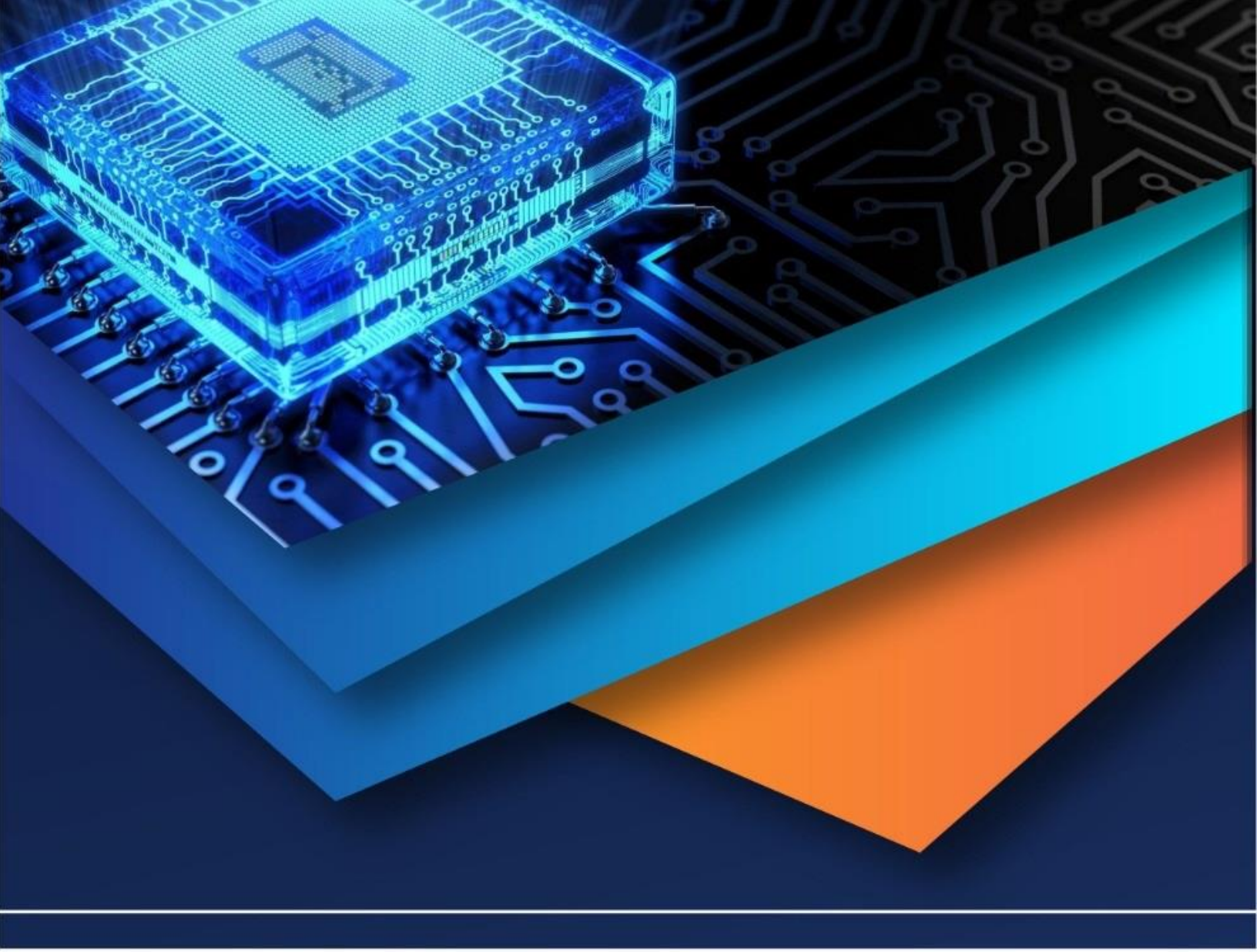
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

played a significant role in shaping the direction and quality of this work. I am also thankful to the faculty of the Centre for AI, MITS Gwalior for providing the necessary resources. I extend my appreciation to my friends for their feedback and support. Finally, I am deeply grateful to my family for their constant motivation and encouragement during the completion of this project.

REFERENCES

[1] J. Weizenbaum, "ELIZA—A Computer Program for the Study of Natural Language Communication," Communications of the ACM, vol. 9, no. 1, pp. 36–45, 1966. Link: https://dl.acm.org/doi/10.1145/365153.365168
[2] A. Abd-alrazaq et al., "Artificial Intelligence-Enabled Chatbots in Mental Health: A Systematic Review," Journal of Medical Internet Research, vol. 22, no. 7, 2020. Link: https://www.jmir.org/2020/7/e16021/
[3] T. B. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot)," JMIR, 2017. Link: https://www.jmir.org/2017/6/e7785/
[4] S. Zhang et al., "A Neural Conversational Model," arXiv preprint, 2015. Link: https://arxiv.org/abs/1506.05869
[5] A. Vaswani et al., "Attention Is All You Need," NeurIPS, 2017. Link: https://arxiv.org/abs/1706.03762
[6] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," NAACL, 2019. Link: https://arxiv.org/abs/1810.04805
[7] T. Brown et al., "Language Models are Few-Shot Learners," NeurIPS, 2020. Link: https://arxiv.org/abs/2005.14165
[8] R. Calvo et al., "Natural Language Processing in Mental Health Applications," IEEE Intelligent Systems, 2017. Link: https://ieeexplore.ieee.org/document/7889648
[9] N. Ghoshal et al., "Chatbot for Mental Health Diagnosis Using NLP and Deep Learning," Springer, 2023. Link: https://link.springer.com/chapter/10.1007/978-981-19-3450-0_41
[10] A. Miner et al., "Talking to Machines About Personal Mental Health Problems," JAMA, 2016.Link: https://jamanetwork.com/journals/jama/fullarticle/2543772
[11] K. Inkster et al., "Digital Mental Health Management Using AI Chatbots," JMIR, 2018. Link: https://www.jmir.org/2018/5/e10164/
[12] S. Shum et al., "Eliza Revisited: Chatbots in Mental Health," AI Magazine, 2018. Link: https://ojs.aaai.org/index.php/aimagazine/article/view/2785
[13] A. Nageswari et al., "AI-Powered Mental Health Chatbot Using NLP and Sentiment Analysis," 2023. Link: https://www.researchgate.net/publication/371918544
[14] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Approach," arXiv, 2019. Link: https://arxiv.org/abs/1907.11692
[15] D. Jurafsky and J. H. Martin, "Speech and Language Processing," 3rd ed., 2023. Link: https://web.stanford.edu/~jurafsky/slp3/
[16] A. Sheth et al., "Cognitive Services and Intelligent Chatbots," IEEE Internet Computing, 2019. Link: https://ieeexplore.ieee.org/document/8667078
[17] M. F. Porter, "An Algorithm for Suffix Stripping," Program Journal, 1980.Link: https://tartarus.org/martin/PorterStemmer/
[18] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, 1997. Link https://www.bioinf.jku.at/publications/older/2604.pdf

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)