International Journal of Advances in
Soft Computing and Intelligent Systems (IJASCIS)
2026, Vol 05, Issue 01, 15-21
ISSN: 3048-4987

Original Paper

MENTAL HEALTH CHATBOT - MINDEASE

Sheetal Pereira[a], Shashank Bhanushali[b]

[a]Assistant Professor, Department of Computer Engineering, K J Somaiya School of Engineering, Mumbai, India.

[b]Student, Department of Computer Engineering, K J Somaiya School of Engineering, Mumbai, India.

Email ids: sheetalpereira@somaiya.edu, shashank07@somaiya.edu

ABSTRACT

Mental health issues are a growing global concern, yet access to timely support remains limited due to stigma, cost, and resource constraints. This paper presents MindEase, an AI-powered mental health chatbot, documenting its evolution from a traditional Feedforward Neural Network (FNN) to an advanced Retrieval-Augmented Generation (RAG) system. We began with a baseline FNN for intent classification, a common chatbot approach. Real-world testing revealed critical limitations: the FNN could only handle narrow, predefined conversation topics, unsuitable for nuanced mental health discussions. This motivated our transition to a RAG system integrating Gemma 2B LLM with a ChromaDB knowledge base covering diverse mental health topics. This paper demonstrates how the RAG approach addresses the conversational breadth and contextual understanding that mental health support demands.

KEYWORDS

Mental Health, AI Chatbot, Natural Language Processing (NLP), FNN, Large Language Model (LLM), Retrieval-Augmented Generation (RAG)

1. INTRODUCTION

Mental health disorders represent a significant global health burden. Access to professional care remains limited by cost, social stigma, and a worldwide shortage of qualified professionals [6]. These challenges are particularly acute among student populations facing high levels of stress, anxiety, and depression [6]. The COVID-19 pandemic has exacerbated these issues, with studies showing a 67.7% increase in suicidal behavior during lockdowns and widespread service disruptions [2].

AI-powered chatbots have emerged as a promising solution [3], [5], [7], offering 24/7, anonymous, low-cost support for individuals hesitant to seek traditional therapy [4], [6]. However, their efficacy depends heavily on the underlying architecture. Many traditional systems rely on rule-based logic or simple intent-classification models [2], [4], [6]. While efficient for predefined topics, these models lack conversational depth and fail when users express complex feelings, leading to rigid, unhelpful responses. To address this, developers began using generative models like Transformers and LLMs [1], which produce human-like, empathetic responses. However, these models risk fabricating information, a critical concern in mental health contexts.

Our Approach: We began with a baseline FNN intent classifier, a proven approach used in existing systems [2], [4], [6]. Real-world testing revealed fundamental limitations: mental health conversations span unpredictable topics that predefined intent structures cannot accommodate. This motivated our development of a RAG-LLM system combining Gemma 2B's conversational flexibility with retrieval-based grounding in a curated knowledge base, enabling nuanced conversations while maintaining factual accuracy.

2. LITERATURE SURVEY

Early mental health chatbots rely on intent classification. The Psykh chatbot uses RASA NLU to determine user emotions and intents [2]. The SAAC chatbot [4] and systems by Prathaban et al. [3] use NLP preprocessing (tokenization, lemmatization, bag-of-words) with sequential ML models. The Maxx chatbot uses DialogFlow for emotional state identification [6]. These systems work well for constrained domains like delivering CBT exercises [7].

However, their primary limitation is a lack of conversational depth. They mimic understanding without true comprehension [2], resulting in a Limited Scope of Support [3]. Maxx performed well for positive emotions but struggled with the variety of problems and unique ways of expressing negative emotions [6]. Traditional systems are task-oriented and do little to understand human social and emotional cues [5].

Newer models use generative AI. CareBot employed a Transformer model (DialoGPT) [1]. Users overwhelmingly preferred its responses for having a more human-like feel and higher empathy [1], suggesting generative models excel at empathetic, open-ended mental health conversations.

However, general-purpose LLMs pose risks, including hallucinations and unsafe responses. Retrieval-Augmented Generation (RAG) addresses this by retrieving information from specialized knowledge bases before generating responses, grounding LLMs in vetted mental health resources and improving safety, transparency, and relevance.

3. System Design and Methodology

MindEase is a web-based application built on a three-tiered architecture (Figure 1). Our research involved designing and evaluating two distinct models.

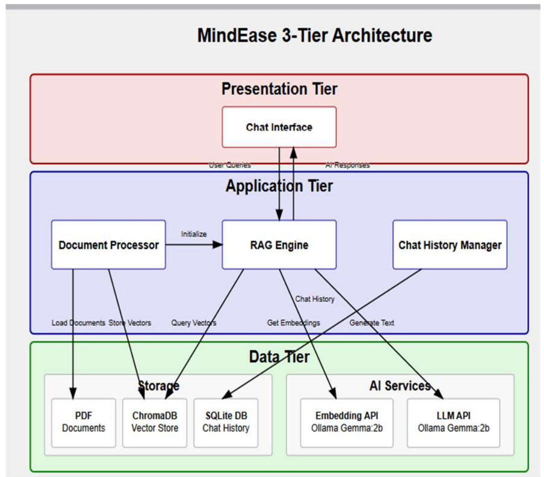

Fig. 1. System architecture of MindEase.

3.1. Baseline Model: FNN Intent Classifier

1. **Architecture:** Feedforward Neural Network built with TensorFlow/Keras
2. **NLP Pipeline:** NLTK-based processing; tokenization, lemmatization, stop-word removal, bag-of-words conversion [3], [4]
3. **Training:** Trained on a manually labeled intents.json file containing tags, patterns, and responses [2], [4], [6]
4. **Function:** Classifies user input into predefined intents and returns hard-coded responses

Figure 2 shows the data flow for intent classification.

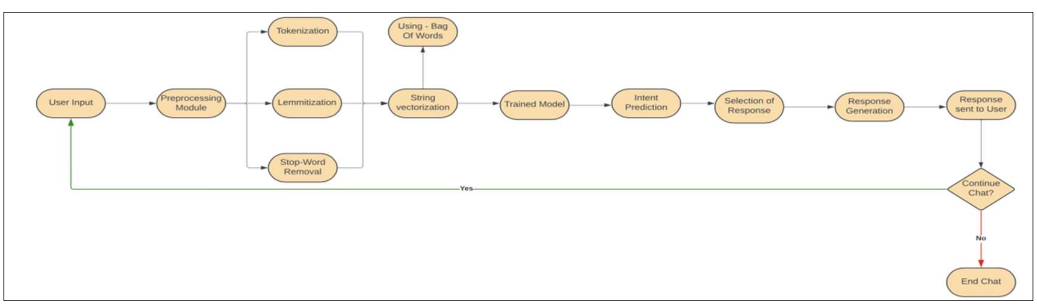

Fig. 2. FNN data flow diagram.

3.2. Advanced Model: RAG-LLM System

After identifying FNN limitations, we developed a RAG architecture for open-ended, context-aware conversations.

System Components:

1. **LLM Generator:** Gemma 2B (via Ollama); lightweight, open-source, produces empathetic responses

2. **Vector Database:** ChromaDB for efficient semantic search
3. **Embedding Model:** Gemma 2B's built-in embeddings
4. **Knowledge Base:** Curated PDFs covering mental health disorders, CBT techniques, mindfulness, coping strategies, crisis resources

RAG Pipeline:

1. **Document Processing:** PDFs split into 700-character chunks (200-character overlap), embedded via Gemma 2B, stored in ChromaDB
2. **Query Processing**: User input converted to embedding, top 3 similar chunks retrieved
3. **Response Generation**: Gemma 2B generates a response (temperature=0.7) from the prompt, combining the query and retrieved context

Optimizations: Response caching, engineered prompts for empathy, temperature=0.7 for coherence/empathy balance, rate limiting (5 req/sec), SQLite for conversation history enabling context-aware multi-turn conversations.

Transition from FNN to RAG:

1. **Scalability**: Mental health coverage requires hundreds of intents; RAG uses an expandable knowledge base
2. **Language Diversity**: Users express concerns variably; semantic search handles linguistic variation
3. **Multi-dimensional Queries**: Mental health spans multiple domains; RAG naturally processes complex queries

4. RESULTS AND DISCUSSION

4.1 Baseline FNN: Performance and Limitations

The FNN achieved >95% accuracy on training data with ~50ms response time. However, practical deployment exposed critical limitations:

Coverage Gap: Trained on around 100 intent categories. Mental health conversations span far beyond exam stress, family conflicts, sleep, relationships, etc. Queries outside predefined intents resulted in "I don't understand" or misclassification.

Example Failures: The startup window is provided in Figure 3.

User: "My girlfriend broke up with me, and now I can't focus on studying."

FNN: "I detect sadness. Try breathing exercises." (Generic, misses complexity)

This confirms that intent-based systems have a limited scope of support [3] and struggle with various unique ways of expressing feelings [6].

4.2 Advanced RAG-LLM: Addressing Limitations

The RAG system addresses FNN limitations through:

1. **Broad Coverage:** Knowledge base covers disorders, therapies, and coping strategies; retrieves relevant information for ANY mental health query.
2. **Contextual Understanding:** LLM synthesizes retrieved information with the user's specific phrasing.

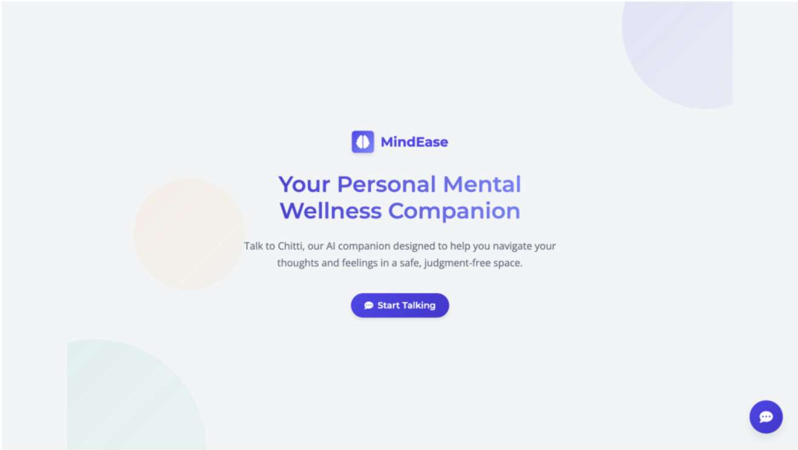

Fig. 3. MindEase Homepage Interface.

Chat Demonstrations (Figures 4-5):

For example, "i was feeling sad today, my exams went bad,"

The system retrieved exam anxiety documents and generated: "I'm so sorry your exams didn't go as planned. It's completely normal to feel sad after a disappointing experience..." with actionable advice.

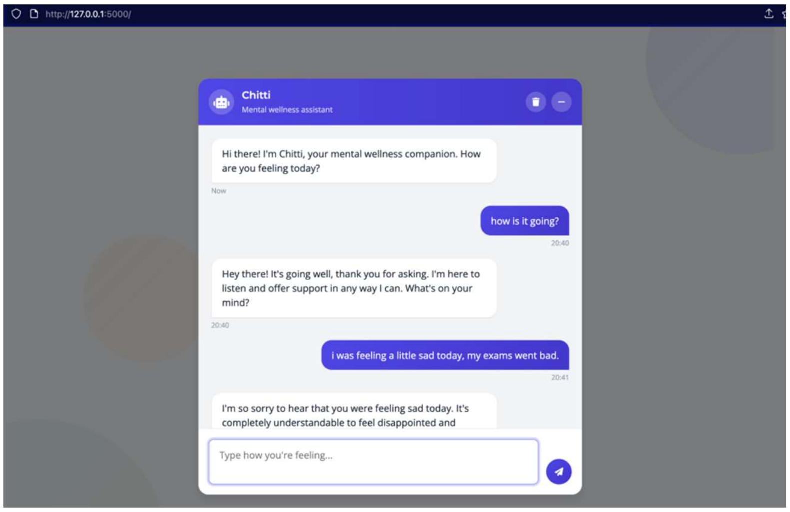

Fig. 4. Chat demo: contextual response.

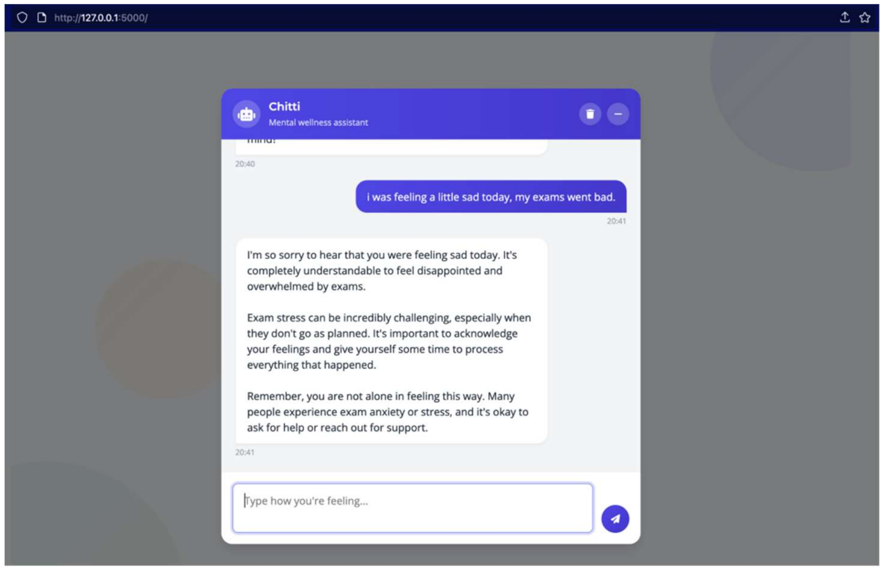

Fig. 5. Chat demo: empathetic engagement.

5. Conclusions

This paper documents MindEase's iterative development from traditional FNN intent classification to the RAG-LLM architecture. Real-world applications revealed that intent-based systems are fundamentally limited for mental health support, requiring conversational breadth and contextual understanding. Our RAG-LLM system, integrating Gemma 2B with a comprehensive knowledge base, successfully addresses FNN limitations, providing flexible, contextually grounded responses across diverse mental health topics.

Key Contributions:

1. Documentation of intent-based approach limitations in mental health chatbots
2. RAG-based alternative with demonstrated superior conversational capability
3. Production-ready architecture with crisis detection and persistent memory

Future Work:

1. User studies measuring satisfaction, engagement, empathy (100+ participants)
2. Sentiment analysis tracking mood progression
3. Multimodal input (voice, images)
4. Hybrid model: FNN for FAQs, RAG for complex queries
5. Multilingual, culturally-adapted knowledge bases
6. Wearable device integration
7. Federated learning preserving privacy
8. Clinical trials measuring effectiveness (PHQ-9, GAD-7)

Authors Contributions (Compulsory):

Ms. Sheetal Pereira: Conceptualization, Supervision, Project Administration, Validation, Writing – Review & Editing. **Shashank Bhanushali**: Conceptualization, Methodology, Software, Data Curation, Writing – Original Draft

REFERENCES

1. Crasto, R., Dias, L., Miranda, D., & Kayande, D. (2021). CareBot: A Mental Health ChatBot. In 2021 2nd International Conference for Emerging Technology (INCET).
2. Patole, A., Dumbre, V., Kesharwani, R., & Khanuja, H. (2021). Mental Health Chatbot (Psykh). International Research Journal of Engineering and Technology (IRJET), 8(3), 1084–1087.
3. Prathaban, B. P., Subash, R., G, L., & A, A. (2024). AI based Mental Health Assisted Chatbot System. In 2024 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS).
4. Jayabhaduri, R., Vijayaraghavan, A., R, A. K., G, C., & Sailesh, S. S. (2024). AI Powered Chatbot for Mental Health Treatment. In 2024 First International Conference on Technological Innovations and Advance Computing (TIACOMP).
5. Jovanović, M., Baez, M., & Casati, F. (2020). Chatbots as Conversational Healthcare Services. IEEE Internet Computing.
6. Dhanasekar, V., Vishali, Y. P., Praveen Joe, S., & Booma Poolan, M. (2021). A Chatbot to promote Students Mental Health through Emotion Recognition. In 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA).
7. Rani, K., Vishnoi, H., & Mishra, M. (2023). A Mental Health Chatbot Delivering Cognitive Behavior Therapy and Remote Health Monitoring Using NLP And AI. In 2023 International Conference on Disruptive Technologies (ICDT).