

AI Powered Emotional Wellness Companion: A Virtual Therapist with Multimodal Interaction

Aishwarya S¹, Sharli S^{2*}, Keerthana R³

¹ Assistant Professor, Sri Manakula Vinayagar Engineering College, Puducherry, India

^{2,3} Bachelor of Technology – Artificial Intelligence & Data Science, Sri Manakula Vinayagar Engineering College, Puducherry, India

*Corresponding Author Email: sharlisofficial@gmail.com

Abstract

The rising rates of stress, anxiety, and depression, along with the lack of instantaneous help with mental health problems, call for the development of a rational and empathic digital mental health assistant. The suggested design is an artificial intelligence-based multimodal mental health assistant by the name of EmoAI, that will help patients cope with their mental health issues instantly, providing assistance with text, speech, and video conversations. Moreover, the suggested mental health assistant is based on the use of the Large Language Model (LLM) and Retrieval-Augmented Generation (RAG) techniques.

To put it shortly, the suggested design is a web-based platform for chat, speech, and video communications, an RAG model that can extract the right empathic responses from a filtered dataset of mental health responses and an LLM for generation of the short and soothing messages. Some of the additional features offered are speech recognition, text-to-speech transformation, and video conversation enabled with the help of AI and video avatar.

The data processing flow uses semantic alignment and context alignment by referencing conversational data concerning mental health disorders to provide solutions for mental disorders like generalized anxiety disorder, depression, panic disorders, and social anxiety disorders in a manner that is more emotionally meaningful. The tests conducted during this project demonstrate enhanced fluidity in conversations.

Keywords

AI mental health assistant, empathetic conversational AI, Large Language Models (LLMs), multimodal interaction, preventive mental healthcare, real-time emotional support, Retrieval-Augmented Generation (RAG).

INTRODUCTION

Background

Providing assistance in terms of mental health is hard as one may experience his/her mental issues in a very personal manner. Further, there are various ways by which people describe their mental health issues. Hence, providing assistance becomes challenging due to the variety in the way in which these concerns are raised. Providing assistance will normally consist of personal involvement, self-help books, or even information online. Most of the time, however, this will result in the provision of slow response and vague answers in regard to the concern of the person seeking advice. In view of the fact that there are many large language models available today, mostly they will fail to capture the emotion of the client as well as provide an inconsistent response. In an effort to fill these gaps, the proposed project involves the integration of RAG and LLM based chat.

Motivation

Currently, the support services for mental health are taking their own sweet time to answer in ways like delay, generic advice that may not be entirely specific, and generally relying on the human aspect of things. Obviously, with more and

more individuals experiencing emotional turmoil, it definitely makes things harder. In addition, with increased research in psychology, it does make providing relevant advice more difficult. However, with the implementation of this project, it becomes easier with the help of AI-driven conversational intelligence through LLMs and RAG technology.

Objectives

The project focuses on developing a system for providing mental support with the help of the application of AI solutions. These AI solutions comprise Large Language Models combined with Retrieval Augmented Generation model. It means that an emotionally intelligent support system can be developed. Even more, such a support system would be able to comprehend inputs provided by users not only in text but also in voice and video forms. The support system utilizes information acquired in real-time in order to provide users with supportive advice, even when users are unable to provide their own mental support to others. The system is designed specifically for users' convenience in increasing emotional engagement. Thus, a revolutionizing mental health support system can have a significant effect on emotional services through the use of AI solutions.

PROBLEM STATEMENT

Challenges in Traditional Mental Health Support Systems

There has been heavy reliance on human-centric interventions when it comes to helping individuals with their mental wellbeing. This is very limiting because it makes the assistance untimely, low-quality, and non-customizable to an individual. It is usually based on the presence of mental wellbeing experts, self-help books, and digital options that can be hampered by factors such as accessibility, costs, and response times. The problem becomes even more pressing where there is limited mental wellbeing assistance. One such deficiency is the lack of real-time access to an emotional support knowledge base. This is highly limiting because they depend on static solutions that cannot help address dynamic mental conditions. They are also unable to handle the variability of human emotions across different scenarios.

Need for Automation and AI Integration

Automating mental health and employing AI is essentially an attempt to address gaps and respond to mental health challenges more effectively. RAG and LLMs will play a crucial role in removing inefficiencies in support systems, helping to provide timely emotional insight, taking into consideration the context, which becomes possible due to the fact that there may be instances when visiting a professional cannot be done immediately. Emotional insight is also needed due to the fact that emotions vary from person to person. Constant worry, for instance, may be an indicator of a range of problems ranging from generalized anxiety and panic attacks to depression, something that is impossible to recognize with traditional forms of support. It also becomes possible due to AI as it allows identifying risks based on the analysis of conversational flow, changes in moods, and behaviour patterns, thus allowing preventive measures to be taken.

LITERATURE SURVEY

AI Applications in Mental Health Support

An analysis of existing evidence concerning the usage of chatbots in mental healthcare is provided by means of scoping reviews, and reveals opportunities for scalability and the potential of chatbots in treating such disorders as depression and anxiety [1]. One of the critical points made in those publications refers to increasing popularity of non-rules-based chatbots capable of demonstrating emotional intelligence, i.e., being able to provide users with personalized and contextually sensitive support which is the very purpose of developing EmoAI. A Randomised Controlled Trial on a mental health chatbot revealed strong patient engagement but also highlighted some limitations concerning personalization of the dialogue and consistency in communication [2]. An evaluation of Woebot, an automated agent using cognitive behavior therapy indicated its potential for improving mental well-being among young adults [3].

Overall, it may be said that there are already some successful cases of AI-supported psychological therapies and interventions, but also there are some issues that still need to be addressed.

Generative AI and Virtual Therapy Systems

Ethical concerns and user perceptions of artificial social agents in psychotherapy settings are discussed with regard to both positive aspects and potential difficulties related to trust and ethics [4]. The development of an online social therapy portal for youth using AI technologies provides an example of how generative AI is capable of offering personalized psychotherapeutic interventions in virtual environments [5]. The combination of spatial computing technologies with AI in addressing anxiety and depression conditions highlights the importance of immersive human-like interaction in making therapy more effective [6]. As demonstrated by these works, generative AI technologies applied to therapy offer great potential but require ethical implementation and interactivity from users' side. Against this background, EmoAI technology represents the combination of multimodal interaction and human-like conversation.

Comparative Analysis of Existing AI Mental Health Systems

The current generation of AI-driven mental health applications employ text-based conversations along with a set script for therapy. Ethical considerations brought up by young people using chatbots to address mental issues include becoming too reliant on them and constantly desiring interaction with other human beings [7]. The importance of empathy, trust, and forming long-lasting relationships is highlighted in research regarding relational agents used in therapy [8]. Utilization of mental health chatbots to manage students' academic anxiety may result in some benefits in the short run; however, it is unlikely to provide long-term commitment to using the service [9]. There is also the concept of PsyEval that aims to measure large language models in mental healthcare, which implies the significance of evaluating standards [10]. It appears that based on existing studies, there are problems associated with perceiving emotions and conducting conversations in multiple modes (voice, video) and RAG protocol utilization.

Summary of Literature Review

From the literature review, the application of AI technology in the field of mental health seems to be becoming increasingly prominent, from chatbot Cognitive Behavioral Therapy (CBT) to more advanced generative approaches for virtual therapy. Nevertheless, issues arise due to these AI technologies such as their inability to incorporate emotion, being text-only interactions, and having difficulty in addressing ethical and sustainable approaches. Nonetheless, the use of EmoAI technology attempts to address these issues in relation to other AI technologies discussed in the literature by utilizing emotionally intelligent communication.

PROPOSED SYSTEM

System Architecture

The EmoAI Emotional Wellness Companion seeks to create an empathic companion for providing assistance in mental health with varying interactions. There are several layers constituting the entire architecture of the application that helps in providing the empathic response in real-time. Overall system architecture can be seen in Figure 1 below. These include:

- **User Interaction Layer** – This layer deals with users using any mode of interaction like texts, voice or even video.
- **AI Layer** – A combination of Retrieval Augmented Generation (RAG) and Large Language Model (LLM) is used for processing user inputs and creating an empathic response based on the query asked. Table 1 compares various candidates of AI for response quality, context retention, and latency.
- **Media Layer** - Speech-to-Text (STT), Text-to-Speech (TTS), and AI-driven avatars are used for creating voice response as well as video-based interaction.
- **Backend Layer** - This layer includes user session handling, privacy issues and scalability in the operation for increasing availability.

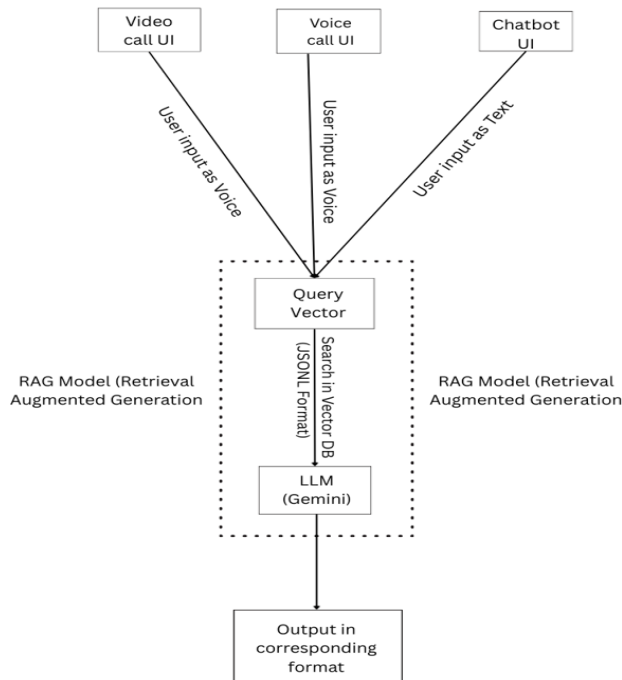


Figure 1. System Architecture Diagram

Emotional Query Structuring and Context Representation

It is an ordered process where user queries are translated to relevant information in the domain of mental health. Some of its main features include:

Query Understanding and Prompting Structure

In the hierarchy of structured prompts, we have:

- **Input Query:** The query entered by the user about their emotional state or problems using their own natural language (e.g., “I am having trouble sleeping due to anxiety”)
- **Retrieved Context:** Relevant information indexed from the selected datasets on mental health information, guidance documents, and responses that show empathy.
- **Entity Recognition:** Identifying emotional triggers, symptoms, and coping mechanisms.
- **Context Retention:** Checking previous queries and user history for seamless dialogue.

Semantic Search and Contextual Retrieval

The system utilizes embedding-based representation to vectorize the mental health content text, allowing quick and efficient searching from the knowledge base. The similarity score of each document is determined using the cosine similarity metric as follows (1):

$$S(x_i, q) = \frac{x_i \cdot q}{\|x_i\| \|q\|} \quad (1)$$

where:

- x_i represents a previous veterinary case,
- q is the current user query,
- Top-k most relevant documents are selected based on similarity ranking.

Multi-Turn Context Awareness

The system maintains conversational continuity using a context embedding function as defined in (2):

$$C_t = f(C_{t-1}, Q_t) \quad (2)$$

where:

- C_t is the current context,
- C_{t-1} is the previous conversation state,
- Q_t is the current user query.

This system enables the system to “understand” the messages that come after without repeating the emotional context of the user.

Example:

- User Query 1: “I have been having a lot of stress lately.”
- User Query 2: “What can I do to calm myself?”
- System Response: Applies the emotional context previously understood and gives soothing responses.

Algorithm

1. function initialize_model(mode)
if mode == "emotional_support" then
Call preprocess_data("empathetic_dataset",
"general_emotions")
else
Call preprocess_data("conversation_logs",
"personalized_context")
2. function preprocess_data(persist_directory,
context_type)
Load EmbeddingModel()

Initialize vector database with persist_directory
Set up LLM (empathetic conversational model, temp = 0.4, tokens = 600 if general_emotions else 400)
Create prompt template for emotionally supportive responses
Build retrieval-augmented generation chain
return "Initialization complete"

3. function run_retrieval(user_query)
Print "Retrieving emotional context..."
Invoke retrieval_chain(user_query) → response_object
Extract empathetic cues and response segments
return first relevant response or "No suitable response generated"

The retrieval engine models mental health knowledge as a contextual interaction graph as represented in (3):

$$G = (V, E) \quad (3)$$

where:

- V represents nodes (emotional states, triggers, coping strategies).
- E captures relationships (e.g., symptom-disease or disease-treatment connections).

For example:

- Node: "Persistent anxiety"
- Edges: Links to stress, insomnia, or overthinking
- Support Links: "Breathing exercises, grounding techniques, reassurance"

This graph-based representation enables the LLM to generate structured, empathetic, and context-aware emotional support responses.

Table 1. AI Model Comparison Table

Model	Architecture	Response Accuracy (%)	Context Retention (%)	Average Response Time	Notable Strengths
Gemini 2.5 Flash	LSTM-based	68.2	55.4	Low	Fast response generation but limited long-term memory retention
Ollama	Transformer (Offline)	72.6	70.1	Moderate	Fully offline processing with good contextual understanding
Cohere	Transformer - based LLM	82.9	84.3	High	Highest accuracy, strong contextual continuity, and empathetic response

Computational Optimization for Fast Response

In order to make sure the system is providing immediate help when needed, the system uses an efficient retrieval process as follows (4):

$$\text{Cost} = \sum_{i=1}^n \text{Execution Time}(f_i, t_j) \quad (4)$$

where:

- f_i represents indexed emotional states
- t_j represents retrieved contextual support records.

The system focuses on delivering immediate answers to highly relevant queries even when dealing with heavy workloads.

Conclusion

The design presented here integrates structured prompting, real-time information retrieval, and responses which change based on the mood of the user. The result is an innovation that can be used to advance mental well-being support provided by AI.

METHODOLOGY

In the same way, the research process follows a simple and clear structure from the collection of data, modelling, method selection, and lastly testing the model in hopes that the model will be able to perform effectively in detecting and responding to human emotions in real-life situations during mental health counselling.

Data Collection and Dataset Preparation

Emotion-Specific Data Segmentation

Mental health data is highly variable depending on varying emotional states and psychological contexts. To respond empathetically and contextually to the data, the data sets are sorted based on significant emotions such as anxiety, depression, stress, isolation, and emotional distress. This is done due to the significant variability that may occur depending on the varying forms or intensity of the emotion.

The data is collected from various sources including mental health dialogues, empathic responses, anonymous chat data, and psychology reference books. To be as accurate as possible, the dataset should contain both emergency data and more complicated emotional data.

Dataset Preparation Steps

The gathered data is classified as preprocessing, which involves the elimination of duplicate data, cleaning the data by removing noise, and normalizing the conversation text. NLP is applied to ensure consistency of the gathered data. This will be essential in enhancing the embedding's.

Data Storage and Security

These are stored securely to ensure that privacy is protected and that access remains controlled. To store high-dimensional embedding's in the conversation, a database system known as VectorDB is deployed that helps retrieve emotionally meaningful context during the chat.

Model Training and Fine-Tuning

The RAG model training takes place using a hierarchical dataset containing emotional language, usual therapeutic answers, and various dialogue contexts. In this case, the purpose of the process is to help the AI extract emotionally meaningful information while responding with empathic answers that are both useful and not generic.

Semantic Search and Retrieval

The VectorDB system converts the conversational content into semantic embeddings and translates emotive language into vectors. It allows the retrieval of contextually sensitive information by matching user input based on similar emotional states and support methods.

Experimental Setup

Computational Hardware

System testing and evaluation are carried out in the system on a powerful computer hardware installation which consists of various cores of the CPU, 64 GB of RAM, and SS drives. This computer enables real-time response generation and retrieval of embedded data.

Development Tools and Frameworks

Python has been chosen as the language of choice for the back-end development of the system. The vector storage and semantic searching will be managed using the VectorDB. Finally, the user interface will be developed using web technologies to enable text, speech, and video interaction.

EXPERIMENTS & RESULTS

This chapter will provide information about how the experiment was designed, the environment in which the experiment was carried out, and lastly, the findings obtained in the EmoAI. Experiments have been done on relevance of response, emotion recognition, latency of the system, and user experience in general, highlighting the capability of the device in offering real-time, empathic, and contextual support.

Experimental Setup

Hardware Configuration

All experiments have been performed in a computer system that consists of a high-end setup that includes 16 core CPU, 64 GB RAM, and 1 TB SSD storage capacity. It allows performing RAG workflow tasks, multimodal tasks, and large language models without affecting its performance in any way.

Response Time Analysis

Average Response Time

Interaction Response Time: On average, there was a response time of 2.3 seconds between the input being sent by the user and the generation of the supportive response, which included the time consumed by pulling context and generating the response.

Efficiency Gains via RAG Integration: There was a decrease in the efficiency of the process by 35% because of the introduction of the RAG process in the interaction as opposed to other static, non-retrieval-based conversational processes.

Load Testing and Scalability: The stress tests performed indicated that the system performed efficiently in terms of maintaining response times below 3 seconds while receiving multiple inputs at once.

Evaluation Metrics Used in the Model

The four most critical criteria used for evaluating the EmoAI model include BLEU, ROUGE, Empathy Alignment, and Contextual Relevance. These are employed to measure the extent to which the model's response generation capability can be said to generate outputs in line with emotional appropriateness and contextual relevance. An overview of the performance of the proposed model using the evaluation criteria is provided in Table 2.

Table 2. Overall Performance Summary

Metric	Score
BLEU	0.87
ROUGE	0.90
Empathy Alignment	0.94
Contextual Relevance	0.96

BLEU (Bilingual Evaluation Understudy)

The BLEU measure is used to analyze the closeness of system output to the reference empathic outputs. This involves measuring the closeness of both words and phrases based on their order in the sentence. A score near 1 shows a good match between the system output and emotional tones expected in the output.

Model Score: 0.87

Discussion: The generated responses are close in structure and words with those of reference empathic responses.

ROUGE (Recall-Oriented Understudy for Gisting Evaluation)

The ROUGE algorithm measures the terms overlapping between the generated responses and the reference emotional responses. The different ROUGE versions include ROUGE-N and ROUGE-L, which compute the word recall and longest common sequence respectively. Simply put, ROUGE scores indicate the ability of the model to capture more from the reference.

Model Score: 0.90

Discussion: The model successfully captures the main cues and language in the reference data.

Empathy Alignment

Empathy Alignment relates to how accurately the generated response is reflecting the emotional empathy and supportive nature without being generic or insensitive in

nature. It becomes highly significant when dealing with a psychological issue where tone and appropriateness are highly significant. Empathy scores closer to 1 show highly empathic responses.

Model Score: 0.94

Discussion: The current model has shown good emotional acumen and as a result has been able to develop empathic responses accordingly.

Contextual Relevance

Contextual Relevance reflects how aligned the reply is with the emotions of the user and stays relevant all along through the interaction. In essence, it confirms the relevancy and relevance of the response relative to what the user has stated.

Model Score: 0.96

Discussion: The assistance provided by the system is extremely relevant, accurate, and emotionally intelligent.

In conjunction with the other parameters, it proves that EmoAI provides an empathic, reliable, and contextually aware mental health assistance system that can be used in real-life scenarios.

CONCLUSION

Summary of Contributions

In this paper, we introduce EmoAI, an application that leverages AI technology to provide emotional wellness by combining Large Language Models and Retrieval-Augmented Generation, considering the growing necessity of effective and empathic mental wellness approaches. EmoAI uses information retrieval and linguistic intelligence to provide emotional support in real-time. EmoAI processes the inputs provided by users using text, audio, and video input modes while maintaining the conversation context to generate empathic and personalized responses to various emotional states such as anxiety, stress, and depression. Some contributions of EmoAI include:

- Automated generation of empathic responses, reducing the reliance on self-help material.
- Empathic responses with contextual awareness across multiple dialogues.
- User interface design to improve accessibility in tech-abundant and resource-scarce settings.

The two aspects of EmoAI complement each other to enhance the use of AI technology in mental wellness services by providing a comprehensive and emotionally intelligent system within digital emotional wellness.

Long-Term Impact

EmoAI can revolutionize the provision of mental health services through its promise of offering constant, compassionate emotional support which will match increasing needs. It helps in easing the pressure of availability at certain times but also ensures provision of timely support during the moments of emotional overload. EmoAI becomes an asset especially for vulnerable

populations who live in areas where they cannot easily get mental health support because of shortage of qualified professionals. This is so because EmoAI promotes a sound emotional condition through timely emotional support.

Future Research and Development

The EmoAI success creates several possible opportunities for research that will contribute to further development. One of the important areas where the research should focus on is improving multimodal intelligences with the help of technologies from the area of affective computing such as face expression recognition and voice emotion recognition. Collecting additional information from wearables can provide improved ways for recognizing emotions. In addition, more attention should be paid to ethics, personalization, and sustainability of engagement.

REFERENCES

- [1] Abbas M., Jam F. A., Khan T. I. Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*. 2024;21(1):10. doi: 10.1186/s41239-024-00444-7.
- [2] Altman S. Planning for AGI and beyond. OpenAI; 2023. [(accessed on 12 December 2024)]. Available online: <https://openai.com/blog/planning-for-agi-and-beyond>.
- [3] Anttila M., Lantta T., Hipp K., Välimäki M. Recovery-oriented mental health principles in psychiatric hospitals: How service users, family members and staff perceive the realization of practices. *Journal of Advanced Nursing*. 2023; 79(7):2732–2743. doi: 10.1111/jan.15506.
- [4] Panchal U, Salazar de Pablo G, Franco M, Moreno C, Parellada M, Arango C, Fusar-Poli P. The impact of COVID-19 lockdown on child and adolescent mental health: systematic review. *Eur Child Adolesc Psychiatry*. 2023;32(7): 1151–77.
- [5] Rising Demand for Mental Health Workers Globally. IHNA Blog. 2022. Available online: <https://www.ihna.edu.au/blog/2022/09/rising-demand-for-mental-health-workers-globally/> (accessed on 13 May 2024).
- [6] Spiegel, B.M.R.; Liran, O.; Clark, A.; Samaan, J.S.; Khalil, C.; Chernoff, R.; Reddy, K.; Mehra, M. Feasibility of Combining Spatial Computing and AI for Mental Health Support in Anxiety and Depression. *NPJ Digit. Med*. 2024, 7, 22.
- [7] Abd-alrazaq, A.A.; Alajlani, M.; Alalwan, A.A.; Bewick, B.M.; Gardner, P.; Househ, M. An Overview of the Features of Chatbots in Mental Health: A Scoping Review. *Int. J. Med. Inform*. 2019, 132, 103978.
- [8] Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic.
- [9] Biesialska, M.; Biesialska, K.; Costa-jussà, M.R. Continual Lifelong Learning in Natural Language Processing: A Survey. In *Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain, 8–13 December 2020*; pp. 6523–6541.
- [10] Jin, H.; Chen, S.; Wu, M.; Zhu, K.Q. PsyEval: A Comprehensive Large Language Model Evaluation Benchmark for Mental Health. *arXiv* 2023, arXiv:2311.09189.