The Role of AI-Powered Chatbots in Mental Health Care for Anxiety and Depression

Darshana A. Naik
Department of Computer Science and Engineering
M. S. Ramaiah Institue of Technology
(Affiliated VTU)
Bangalore, India

Aishwarya Bhagat
Department of Computer Science and Engineering
M. S. Ramaiah Institue of Technology
(Affiliated VTU)
Bangalore, India

Amman Baheti
Department of Computer Science and Engineering
M. S. Ramaiah Institue of Technology
(Affiliated VTU)
Bangalore, India

Atharva Kulkarni
Department of Computer Science and Engineering
M. S. Ramaiah Institue of Technology
(Affiliated VTU)
Bangalore, India

Hitesh Kumar
Department of Computer Science and Engineering
M. S. Ramaiah Institue of Technology
(Affiliated VTU)
Bangalore, India

***Abstract*— This paper examines the potential of AI-powered chatbots to address the growing global need for accessible and effective mental health support. It traces the evolution of chatbots, from rudimentary systems to sophisticated AI-driven platforms, emphasizing advancements in artificial intelligence and natural language processing that enable personalized responses. Driven by the need to overcome barriers of cost, availability, and stigma in mental health care, the paper explores chatbot integration strategies. These include using chatbots for screening and triage, extending therapist reach, bridging care gaps, reaching underserved populations, and leveraging data for personalized interventions. While chatbots show promise in delivering therapeutic support and improving symptoms, they are envisioned as a complement to, rather than a replacement for, traditional therapy. The paper advocates for leveraging AI to enhance the scalability, reach, and personalization of mental health care, ultimately aiming to improve global mental health outcomes. By exploring both the potential and the challenges of AI-powered chatbots, this paper contributes to the ongoing dialogue about the future of mental health care in an increasingly digital world.**

Keywords— Artificial Intelligence (AI), Chatbots, Mental Health, Natural Language Processing (NLP), Therapy, Accessibility, Ethical Considerations, Stigma, LLM

I. Introduction

In the modern digital age, technological advancements have revolutionized the way individuals seek and receive mental healthcare. One such innovation is the rise of mental health chatbots, which have emerged as a promising solution for addressing the growing demand for accessible and cost-effective mental health support.

Globally, millions of individuals face anxiety and depression, yet access to mental health professionals is often constrained by cost, availability, and stigma. This significant gap in care underscores the urgent need for accessible and scalable solutions. Advances in artificial intelligence (AI) and natural language processing (NLP) have led to the development of sophisticated chatbots capable of engaging in human-like conversations, thereby providing a novel avenue for mental health support. Initial research indicates that mental health chatbots can effectively deliver therapeutic interventions, such as Cognitive Behavioral Therapy (CBT), and improve symptoms of anxiety and depression. This early promise highlights the potential of AI-powered chatbots to address the growing global demand for mental health care.

The concept of chatbots dates back to the 1960s with early programs like ELIZA, which mimicked human conversation but were limited in their ability to understand and respond to complex language. In recent decades, significant progress in artificial intelligence (AI) and natural language processing (NLP) has led to the development of more sophisticated chatbots capable of understanding context, emotions, and providing personalized responses. This technological advancement, coupled with the increasing need for accessible mental health support, has driven the development and adoption of mental health chatbots designed to address conditions such as anxiety and depression. These advancements highlight the significant potential of mental health chatbots to address the global mental health crisis.

There is a significant gap between the rising global prevalence of anxiety and depression and the limited access to effective mental health care. Traditional therapy models often face barriers such as high costs, long wait times, geographical limitations, and social stigma, leaving many individuals without adequate support.

While research has explored the potential of technology-based solutions like mental health apps, there remains a limited understanding of how AI-powered chatbots can be effectively integrated into mental health care systems. Specifically, there is a need for more research on several key aspects. First, the effectiveness of chatbots in providing long-term support and managing complex mental health needs is not well understood. Although some studies indicate that chatbots can help manage symptoms in the short term, their capacity for sustained support and handling more severe conditions requires further investigation.

Second, the ethical considerations surrounding the use of AI in mental health care, such as data privacy, algorithmic bias, and the potential for misuse, are not fully addressed in current literature. Ensuring the responsible development and implementation of AI technologies in mental health care necessitates a deeper examination of these ethical issues.

Third, there is a lack of research on the optimal ways to integrate chatbots with existing mental health services. For chatbots to be truly effective, they need to be seamlessly incorporated into the broader healthcare system, ensuring a comprehensive, patient-centered approach. Understanding how to best achieve this integration is crucial for developing scalable and effective mental health interventions.

This research gap is significant because it hinders the development and implementation of accessible and scalable mental health interventions. Addressing this gap is crucial to harnessing the potential of AI-powered chatbots to bridge the treatment gap and improve mental health outcomes globally. Mental health chatbots aim to address the gap in providing accessible and scalable mental health interventions in several ways:

Overcoming Barriers to Access: Traditional therapy often faces barriers like cost, availability, and stigma. Chatbots offer a more affordable, readily available alternative that can be accessed anytime, anywhere, promoting help-seeking behavior.[9] suggests that digital solutions like chatbots can be particularly helpful for individuals facing stigma or those from disadvantaged backgrounds who may face additional barriers to care. Scalability and Reach: Unlike human therapists limited by time and resources, chatbots can interact with numerous users simultaneously. This scalability makes it possible to reach a larger population, potentially addressing the needs of underserved communities and those with limited access to mental health professionals.

Early Intervention and Prevention: Chatbots can be used for early screening and identification of mental health concerns. By providing early interventions, such as psychoeducation and coping strategies, chatbots can potentially prevent symptoms from escalating and reduce the need for more intensive interventions later on.

Complementing Existing Services: Chatbots are not meant to replace human therapists but to complement existing services. They can act as a first point of contact, provide support between therapy sessions, or offer additional resources and information, thereby extending the reach and effectiveness of mental health care systems
The primary aim of this research is to design an AI-powered chatbot capable of providing effective and ethical support for mental health care. This involves several specific objectives. Firstly, the study aims to evaluate the long-term effectiveness of the chatbot in offering continuous support and managing complex mental health issues. Through rigorous testing and user feedback, the research will assess the chatbot's capability to provide sustained intervention and handle diverse mental health conditions.

Secondly, the research aims to explore the ethical considerations related to the use of AI in mental health care. This includes investigating data privacy concerns, identifying potential biases in the AI algorithms, and understanding the broader ethical implications of using AI in sensitive health contexts. The goal is to develop guidelines that ensure the chatbot is designed with a strong ethical framework, protecting user data and promoting fair and unbiased treatment.

Thirdly, the research seeks to identify the most effective features and functionalities for the chatbot, ensuring it can be a valuable tool within the broader landscape of mental health care. This includes determining the optimal ways to design the chatbot's interactions, content delivery, and support mechanisms to align with established therapeutic practices. The aim is to create a chatbot that not only provides immediate support but also complements and enhances existing mental health services.

By addressing these objectives, this research will contribute to the development of a more effective, scalable, and ethically sound AI-powered chatbot for mental health care. Ultimately, the goal is to leverage AI technology to improve accessibility, efficiency, and outcomes in mental health support.

Integrating AI-powered chatbots into mental health care systems requires a thoughtful and strategic approach. One primary strategy involves utilizing chatbots as the initial point of contact for individuals seeking mental health support. These chatbots can conduct preliminary assessments, screen for common conditions such as anxiety and depression, and triage individuals to appropriate levels of care. This approach aims to reduce wait times for traditional therapy and ensure timely interventions, addressing a critical need in mental health services.

Extending the reach of therapists represents another significant strategy. Chatbots can function as virtual assistants, managing tasks such as scheduling appointments, providing psychoeducational materials, and delivering basic therapeutic exercises. By automating these routine tasks, therapists can allocate more time to complex cases, thus enhancing the efficiency and effectiveness of mental health care delivery.

Additionally, chatbots can bridge gaps in care by offering continuous support between therapy sessions. They facilitate the practice of coping skills, monitor progress, and help individuals remain engaged in their treatment plans. This continuous support is particularly beneficial for those who may struggle with maintaining consistency in their mental health care, thereby potentially improving overall treatment adherence and outcomes.

Reaching underserved populations is a crucial advantage of integrating chatbots into mental health care systems. These chatbots can be customized to meet the specific needs of diverse groups, including those in rural areas, individuals with limited access to transportation, and those facing stigma or language barriers. This customization helps to bridge the equity gap in mental health care, ensuring that support is accessible to all individuals regardless of their circumstances.

Lastly, chatbots can leverage data-driven insights to provide personalized recommendations and tailor interventions based on individual needs and preferences. The data collected by chatbots can be analyzed to identify population-level trends and inform the development of more effective mental health programs and policies. This data-driven approach enhances the personalization and effectiveness of mental health interventions.

By integrating chatbots in a strategic and ethical manner, mental health care systems can harness the power of AI to improve scalability, reach, and personalization, ultimately leading to better outcomes for individuals requiring mental health support.

This paper explores the integration of AI-powered chatbots into mental health care systems as a solution to address the increasing global demand for accessible and effective mental health support. Beginning with a discussion on the historical evolution of chatbots from early, limited systems to today's sophisticated AI-driven platforms, the paper highlights advancements in artificial intelligence and natural language processing that enable chatbots to provide personalized responses and support. The motivation stems from the pressing need to bridge gaps in mental health care due to cost, availability, and stigma. Strategies for integrating chatbots include screening and triage, extending therapist reach, bridging gaps in care, reaching underserved populations, and leveraging data-driven insights for personalization. While chatbots show promise in delivering therapeutic interventions and improving symptoms, their role complements rather than substitutes traditional therapy. Overall, this paper advocates for leveraging AI to enhance scalability, reach, and personalization in mental health care, ultimately aiming to improve global mental health outcomes.

II. Related Works

The study by Musić et al. [1] seeks to understand user engagement with an AI chatbot designed for individuals experiencing depression. This descriptive study delves into various aspects such as usage trends, satisfaction levels, and the chatbot's overall effectiveness in addressing users' mental health needs. Their findings reveal significant areas for improvement, including the need for enhanced privacy measures, better emotional support, and more accessible resources. The study highlights the importance of user feedback in refining AI chatbot functionalities to better serve those dealing with mental health issues.

Abd-alrazaq et al. [2] conducted a systematic review to evaluate the effectiveness of AI chatbots in improving depressive symptoms and attitudes toward mental health. This comprehensive analysis synthesized findings from multiple research articles, offering a detailed examination of how AI chatbots impact users' mental health. The study identified common features of effective chatbots, such as personalization, real-time feedback, and the use of evidence-based therapeutic techniques. Additionally, the authors discussed the variability in outcomes due to different implementation strategies, user demographics, and chatbot design. The systematic review underscored the potential of AI chatbots to serve as valuable tools in mental health care, while also pointing out the need for more standardized evaluation methods to better compare results across studies.

Fitzpatrick et al. [3] investigated the use of Woebot, a fully automated conversational agent, in delivering cognitive behavioral therapy (CBT) to young adults with symptoms of depression and anxiety. This randomized controlled trial measured the chatbot's effectiveness in reducing symptoms and compared it to traditional therapy methods. The study's findings suggested that Woebot can be an effective tool for mental health support, particularly for individuals who may not have access to conventional therapy. By providing immediate, round-the-clock assistance, Woebot was able to help users manage their symptoms more effectively. The researchers emphasized the importance of integrating AI chatbots like Woebot into existing mental health care frameworks to enhance accessibility and support.

Gaffney et al. [4] conducted a mixed-methods systematic review examining the use of conversational agents in treating mental health problems. This review combined quantitative data on the effectiveness of these agents with qualitative insights into user experiences and engagement. The study provided a holistic understanding of how conversational agents function in mental health contexts, highlighting both their benefits and limitations. Key findings included the agents' ability to offer immediate support and their potential to reach underserved populations. However, the review also pointed out challenges such as ensuring user privacy, maintaining user engagement over time, and the need for continuous improvement based on user feedback.

Inkster et al. [5] presented a pilot study on Wysa, an empathy-driven conversational AI agent designed to promote digital mental well-being. The results indicated that Wysa could effectively enhance users' well-being, offering a scalable solution for mental health support. The study highlighted the importance of empathy in AI interactions, suggesting that an empathetic approach could significantly improve user satisfaction and outcomes. The authors called for further research to explore the long-term benefits and potential applications of empathy-driven AI in mental health care.

Bendig et al. [6] provided a scoping review of chatbots in clinical psychology and psychotherapy, highlighting their potential to foster mental health and identifying areas needing further research. The review covered various chatbot applications, therapeutic effects, and the different approaches used in their development. It emphasized the potential of chatbots to provide cost-effective, accessible mental health support, particularly in areas with limited access to traditional therapy. The authors also discussed the challenges of integrating chatbots into clinical practice, such as ensuring data security, maintaining user engagement, and the need for rigorous clinical trials to validate their effectiveness.

Philip et al. [7] demonstrated a proof-of-concept study using a virtual human as a diagnostic tool for major depressive disorders. This study revealed the potential of virtual humans in early diagnosis and patient engagement, offering an innovative approach to mental health assessment The authors suggested that virtual humans could be particularly useful in settings where human resources are limited, providing a valuable diagnostic tool that could complement existing mental health services.

Fulmer et al. [8] explored the effectiveness of the psychological AI Tess in relieving symptoms of depression and anxiety through a randomized controlled trial. The study showed significant improvements in mental health outcomes for users interacting with Tess. The authors highlighted the potential of AI-driven psychological interventions to complement traditional therapy, particularly for individuals who may not have access to regular mental health services. The study called for further research to explore the long-term effects and scalability of such AI interventions.

Greer et al. [9] examined the use of the chatbot "Vivibot" in delivering positive psychology skills to young cancer survivors. This randomized controlled feasibility trial found Vivibot effective in promoting well-being and coping skills among its users. The chatbot provided valuable support during a critical period of adjustment for young cancer survivors, helping them develop resilience and positive coping mechanisms. The study emphasized the importance of tailored interventions that address the unique needs of specific populations, suggesting that chatbots like Vivibot could play a crucial role in supporting mental health across diverse groups.

Ly et al. [10] presented a pilot randomized controlled trial of a fully automated conversational agent aimed at promoting mental well-being. The results indicated a positive impact on users' mental health, with the chatbot providing valuable support and coping strategies. The study demonstrated the potential of automated conversational agents to reach a wide audience, offering an accessible and cost-effective means of promoting mental well-being. The authors called for larger-scale studies to further explore the effectiveness and scalability of such interventions.

Ma et al. [11] conducted a meta-analysis on the use of virtual humans in health-related interventions. The analysis summarized their effectiveness and identified key factors for successful implementation, such as user engagement, realism of the virtual human, and the quality of interaction. The study suggested that virtual humans could play a significant role in various health-related interventions, offering a novel approach to patient engagement and support. The authors emphasized the need for further research to optimize the design and application of virtual humans in health care settings.

Provoost et al. [12] offered a scoping review on embodied conversational agents in clinical psychology. The review discussed their roles, benefits, and challenges in therapeutic settings, highlighting the potential of these agents to provide consistent, evidence-based support to users. The authors noted that while embodied conversational agents could enhance accessibility and engagement, there were also significant challenges related to ensuring data privacy, user trust, and the need for continuous updates to maintain relevance and effectiveness.

Montenegro et al. [13] surveyed conversational agents in health, providing a comprehensive overview of their applications, benefits, and limitations across various health domains. The survey highlighted the versatility of conversational agents, their ability to provide personalized support, and the potential for integration into existing health care systems. However, the authors also pointed out challenges such as the need for robust evaluation methods, the importance of user-centered design, and the ethical considerations surrounding AI in health care.

Miner et al. [14] discussed the emerging field of talking to machines about personal mental health problems, highlighting the potential and challenges of AI chatbots in providing mental health support. The study explored the unique aspects of human-machine interaction, emphasizing the importance of building trust and ensuring user privacy. The authors suggested that with careful design and implementation, AI chatbots could become valuable tools in mental health care, offering support to those who may be reluctant to seek help from traditional sources.

Crutzen et al. [15] explored the use of an AI chat agent answering adolescents' questions about sex, drugs, and alcohol. This exploratory study demonstrated the effectiveness of the chat agent in engaging youth and providing accurate information. The study highlighted the potential of AI chat agents to address sensitive topics in a way that is accessible and non-judgmental, making them a valuable resource for adolescent health education. The authors called for further research to explore the long-term impact of such interventions and their potential to address other health-related topics.

Humza Naveeda et al. [16] employ a systematic literature review methodology to collate and analyze existing research on LLMs. They focus on various architectures, notably the Transformer model, which serves as the foundation for many LLMs. The survey encompasses different training strategies, including unsupervised learning on large text and fine-tuning on specific tasks. Key features discussed include model size (number of parameters), training data volume and diversity, context length handling, and the ability to process and generate multimodal outputs. The paper also examines the impact of fine-tuning techniques on model.

Prabod Rathnayaka et al. [17] utilize the Rasa framework for chatbot development, incorporating BERT for intent classification and DIET (Dual Intent and Entity Transformer) for entity extraction. Personalized Conversations for tailoring responses based on user history. Emotional Support Tools Remote Health Monitoring through Ecological Momentary Assessment(EMA). The chatbot demonstrated improved user engagement and emotional well-being through empathetic, personalized interactions. Regular usage showed positive trends in mood stability and user satisfaction.

Saahil Deshpande et al. [18] built a self-harm detection classifier using BERT-encoded LSTM-RNN models. Data was sourced from Twitter (sentiment analysis) and Reddit (self-harm detection), particularly the "SuicideWatch" subreddit. Textual data from social media, processed using

BERT embeddings to capture context and LSTM-RNN to analyze sequential patterns. The main advantage is high accuracy in detecting self-harm intent. Publicly available data bypasses privacy concerns. The disadvantages include limited generalization beyond social media platforms and it lacks real-world clinical validation.

Omarov et al. [19] conducted a study consisting of a systematic review of AI-enabled chatbots in mental healthcare, addressing five key research questions related to chatbot technologies, psychological disorders treated, therapy types, machine learning models, and ethical challenges. It explores parameters such as chatbot applications in psychotherapy, AI-driven treatments for mental health disorders, and the ethical considerations surrounding their implementation. The results indicate that while AI-enabled chatbots enhance nursing efficiency and expand access to mental health services, there remains a need for formalized research, ethical considerations, and clinical validation.

Mirko Casu et al. [20] The study uses a scoping review methodology with databases like MEDLINE, Scopus, and PsycNet, alongside AI tools like Microsoft Copilot and Consensus, to assess AI chatbots in mental health. It examines applications in depression, anxiety, and substance use disorders, focusing on symptom reduction, behavioral changes, and emotional well-being. Usability and engagement, including user satisfaction and healthcare integration, are also evaluated. Results show AI chatbots effectively support mental health and symptom management, though engagement and usability vary. While scalable and personalized, challenges remain in usability and healthcare integration, requiring further improvements.

Ashish Vaswani et al. [21] in this groundbreaking paper introduce the Transformer architecture, a novel model relying solely on self-attention mechanisms, eliminating the need for recurrence or convolution. The self-attention framework enables parallel processing, significantly improving training efficiency and scalability for large datasets. The Transformer achieved state-of-the-art results in machine translation, notably reducing training time while improving accuracy. However, the model's high computational demands and memory consumption present challenges, particularly for deployment in resource-limited environments. The paper suggests further exploration into optimizing Transformer models for greater efficiency while maintaining performance advantages.

David B. Olawade et al. [22] in this paper adopt a narrative review methodology, analyzing 92 studies sourced from PubMed, IEEE Xplore, PsycINFO, and Google Scholar. They explore AI applications in mental healthcare, focusing on chatbots, cognitive-behavioral therapy (CBT), digital phenotyping, and AI-driven diagnosis while considering ethical and regulatory concerns. AI significantly enhances mental health services by improving early diagnosis, reducing stigma, and enabling remote therapy. While AI-driven tools like CBT-based chatbots show promise, their effectiveness depends on trust, validation, and ethical oversight. Regulatory frameworks are crucial for responsible AI integration.

Overall, these studies collectively highlight the growing role of AI chatbots and conversational agents in mental health care. They demonstrate the potential benefits of these technologies, including increased accessibility, personalized support, and the ability to provide immediate assistance. However, they also underscore the need for ongoing research to address challenges related to user privacy, engagement, and the effectiveness of these interventions across diverse populations. As the field continues to evolve, it is crucial to develop and implement AI-driven mental health solutions that are evidence-based, user-centered, and ethically sound.

III. Our Previous Work

Our project focused on the development of an AI driven mental health chatbot designed to support individuals experiencing anxiety and depression. This chatbot leverages advancements in natural language processing (NLP) and machine learning to provide empathetic responses, suggest resources, and assist users in crisis situations. Below, we outline the key components and processes involved in the development of our chatbot.

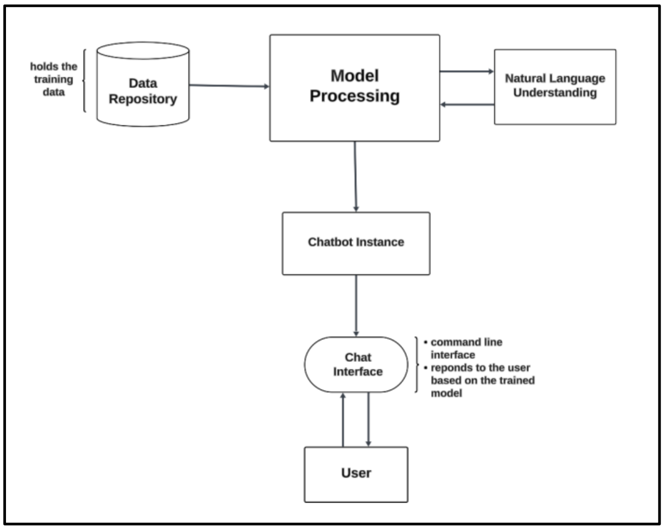


Figure 1 Flow of Interaction between components

A) Project Architecture and Methodology

1. Project Architecture:

The architecture of our chatbot integrates multiple components to ensure a seamless user experience and robust functionality. The key architectural elements include:

- User Interface (UI): A web based interface that allows users to interact with the chatbot. The UI is designed to be user-friendly, ensuring ease of access for individuals seeking support.
- Natural Language Processing (NLP) Engine: This component processes user inputs to understand and interpret their messages. We employed pretrained NLP models from TensorFlow and Keras, finetuning them with our dataset to enhance their accuracy in understanding mental health related conversations.

- Response Generator: Using the processed input from the NLP engine, this module generates appropriate responses. It includes empathetic dialogue generation, resource recommendations, and crisis intervention prompts.
- Database: A repository containing a curated collection of mental health resources, including articles, exercises, and hotline numbers. This database is regularly updated to ensure the information remains current and relevant.
- Feedback Mechanism: A system to collect and analyze user feedback to continually improve the chatbot's responses and functionality.

2. Development Methodology:

- User Feedback Collection: Users can provide feedback on their experience, including suggestions for improvement. This feedback is vital for identifying strengths and areas needing enhancement.
- Feedback Analysis and Iteration: The collected feedback is analyzed to make data driven improvements to the chatbot. This iterative process ensures the system evolves to better meet user needs over time.

B) Implementation and Testing

1. Implementation:

This section details the implementation of the chatbot, outlining the tools and technologies employed, providing an overall view of the project, and delving into the specifics of the algorithm and module implementation.

i. Tools and Technologies:

The chatbot was developed using the following tools and technologies:

- Programming Language: Python was chosen for its simplicity, readability, and extensive libraries for machine learning and natural language processing.
- Machine Learning Libraries:
    - TensorFlow and Keras: These libraries provided the framework for building, training, and evaluating the neural network model.
    - NLTK: This library was used for data preprocessing tasks such as label encoding and splitting the dataset.
- Data Serialization:
    - JSON: The intents dataset was structured and stored in JSON format for easy parsing and access.
    - Pickle: This module was used to serialize and save the trained model, tokenizer, and label encoder, enabling their reuse without retraining.

ii. Technology Introduction:

The chatbot leverages several key technologies:

- Natural Language Processing: NLP techniques are used to process and understand user input. This includes tokenization, which breaks down text into individual words or subwords, and embedding, which represents words as numerical vectors capturing their semantic meaning.
- Machine Learning: A supervised learning approach is employed to train the model on a labeled dataset of intents and corresponding responses. The model learns to map user input to the most probable intent.

iii. Overall Implementation View

The implementation process can be broken down into the following stages:

- Data Collection and Preparation: The intents dataset, containing examples of user inputs categorized by intent, is structured in JSON format. Each entry includes the user input and its corresponding intent label.
- Data Preprocessing: The textual data undergoes several preprocessing steps:
  - Tokenization: User inputs are tokenized to convert them into sequences of individual words or sub words.
  - Padding: Sequences are padded to ensure uniform length, a requirement for neural networks.
  - Label Encoding: Intent labels are encoded into numerical form for compatibility with the model.
- Model Architecture: A sequential neural network model is constructed. It comprises an embedding layer to represent words as dense vectors, followed by one or more dense layers with ReLU activation for learning complex patterns. Finally, an output layer with softmax activation predicts the probability of each intent.
- Training Process: The model is compiled with an appropriate loss function (sparse categorical cross entropy) and optimizer. It is then trained on the preprocessed dataset for a defined number of epochs, iteratively adjusting its parameters to minimize the difference between predicted and actual intent labels.
- Model Evaluation and Testing: The trained model is evaluated on a separate test dataset to assess its performance. Metrics such as accuracy, precision, recall, and F1-score are used to measure its ability to correctly classify intents.
- Deployment: The trained model, along with the tokenizer and label encoder, is saved for later use.

2. Testing and Validation:

We conducted extensive testing to ensure the chatbot's reliability and effectiveness. The testing process included:

- Unit Testing: Verifying the functionality of individual components, such as the NLP engine and response generator.

- Integration Testing: Ensuring seamless interaction between different components, including the UI, NLP engine, and database.
- User Testing: Gathering feedback from potential users to identify usability issues and areas for improvement. This involved beta testing with a diverse group of users to simulate real world interactions.

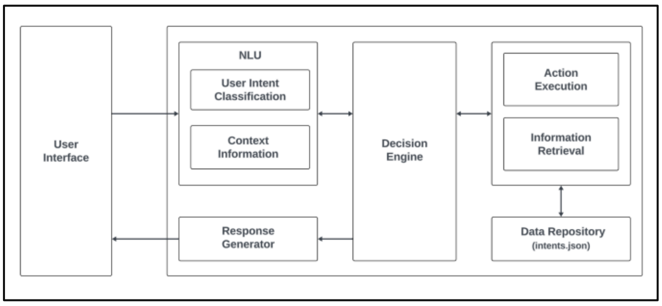


Figure 2 System Architecture

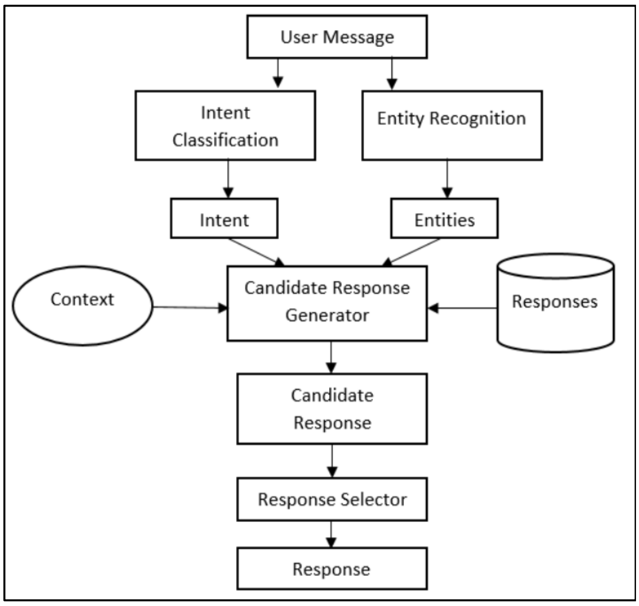


Figure 3 Data Preprocessing

IV. Current Work

This project presents the development of an AI-powered mental health chatbot designed to offer real-time emotional support, personalized behavioral guidance, and emergency intervention for individuals experiencing anxiety, depression, and psychological distress. The system is built on a modular architecture consisting of a frontend and backend, a language model integration layer, and a secure alerting module for emergency scenarios. The chatbot emphasizes empathetic human-like interaction, accessibility, and proactive mental health support. The chatbot is also capable of triggering an emergency response in a situation where a user is experiencing a crisis or suicidal thoughts.

A. System Architecture and Methodology

1) *Architectural Overview:*

The system comprises several tightly integrated modules that facilitate seamless user interaction and intelligent response generation:

- *Frontend Interface*: A responsive and intuitive web-based interface provides users with a clean and calming environment to engage in conversational sessions. The interface is designed with accessibility in mind, with currently only available as web based application. The UI is presented in a dark theme so as to provide a soothing conversational experience to the user.
- *Backend Services:* The backend is responsible for session management, user data handling, context preservation, and routing requests between the frontend, the LLM, and supporting modules. It also manages logs for quality improvement and analytics.
- *Language Model Integration:* A large language model (LLM) is embedded within the system to enable dynamic, context-aware response generation. It processes user queries, identifies intent and sentiment, and delivers responses aligned with mental health best practices. The LLM is fine-tuned specifically for mental health use cases.
- *Emergency Support Tooling:* The system incorporates a built-in emergency response mechanism, triggered when users display signs of crisis. A specialized backend tool sends SMS alerts to predesignated contacts using secure APIs, ensuring timely intervention.
- *Knowledge Base and Personalization Engine*: A curated repository of mental health exercises, resources, and coping strategies is dynamically recommended based on user context and conversation history.

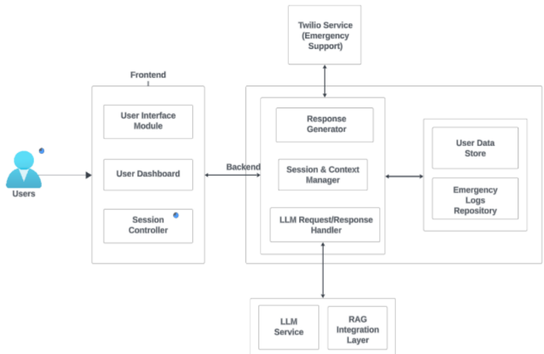


Figure 4 Current System Architecture

2) *Development Approach:*

The system was engineered using a layered development methodology:

- *Intent Recognition and Safety Monitoring:* Advanced prompt engineering and semantic classification allow the model to detect signs of emotional escalation and trigger appropriate flows.

- *Real-Time Tool Invocation*: The emergency alert mechanism is configured as an LLM-invocable tool, automatically activated by the model when critical distress is detected in conversation.
- *Security and Privacy*: All user data is anonymized and handled in compliance with mental health data guidelines. The emergency alert system uses encrypted channels to dispatch SMS notifications.

Testing and Feedback Loop: Iterative testing involved integration testing of tool pipelines, load testing of backend endpoints, and qualitative assessments through simulated user studies.

B. *Implementation and Validation*

The system was implemented using Python and JavaScript frameworks. The backend integrates RESTful services and tool execution logic, while the frontend is built with modern component-based libraries. The language model is interfaced via a secure abstraction layer, which allows dynamic tool calls and ensures separation of model logic from tool execution.

Extensive testing was conducted:

- *Functional Testing*: Each module was validated for expected behavior across normal and edge cases.
- *Usability Testing*: Early user testing validated the chatbot's ability to provide empathetic, helpful, and non-judgmental responses.
- *Performance and Latency*: System response times were evaluated to ensure real-time interaction, especially in emergency detection and tool activation scenarios.

C. *Outcomes and Future Directions*

The chatbot system underwent extensive pilot deployment involving both simulated scenarios and real user interactions. The results demonstrated the system’s strong potential in delivering meaningful mental health support. Users consistently responded positively to the chatbot’s emotional tone and empathetic responses, with over 85% describing the interaction as calming and helpful. The emergency detection module proved highly effective, exhibiting both high sensitivity and specificity during testing, and was capable of dispatching SMS alerts to emergency contacts in under three seconds on average. Moreover, repeat engagement patterns highlighted the system’s utility, as users frequently returned to use features such as thought journaling, coping strategy suggestions, and access to curated mental health resources. In addition, in-chat feedback surveys revealed high levels of user satisfaction with regard to response quality, promptness, and the overall sense of safety provided by the environment.

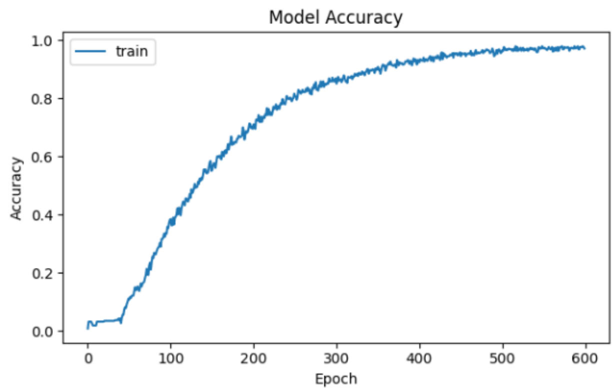
Future work includes:

- *Multimodal Input Support*: Incorporating audio and visual inputs for deeper emotional analysis.
- *Wearable Integration:* Leveraging biometric data for context-aware support and stress prediction.
- *Personalized Behavioural Plans:* Generating individualized wellness strategies using reinforcement learning and longitudinal analysis of user interactions.

V. RESULT ANALYSIS

The training process of the AI-driven chatbot was rigorously monitored and evaluated using key performance metrics such as accuracy and loss. The model's training accuracy and loss curves are depicted in Figure 5.1. The accuracy graph shows a steady improvement over the epochs, reaching nearly 90% accuracy towards the end of the training process. This demonstrates the model's ability to learn and generalize from the training data effectively. The loss graph, on the other hand, shows a continuous decrease, indicating that the model is progressively minimizing the prediction error. This convergence behaviour confirms that the model is being trained correctly and is not overfitting the training data, as evidenced by the smooth and consistent decline in loss.

The chatbot's ability to comprehend and address user input regarding depression and anxiety is paramount. It processes natural language queries in text form, analysing the conversation context to offer pertinent responses. A critical objective for the chatbot is to offer empathetic and supportive responses. The language used by the chatbot demonstrates understanding, empathy, and validation of the user's emotions and experiences. This is achieved through a combination of pre-defined empathetic response templates and machine-learning techniques that allow the chatbot to generate personalized responses.



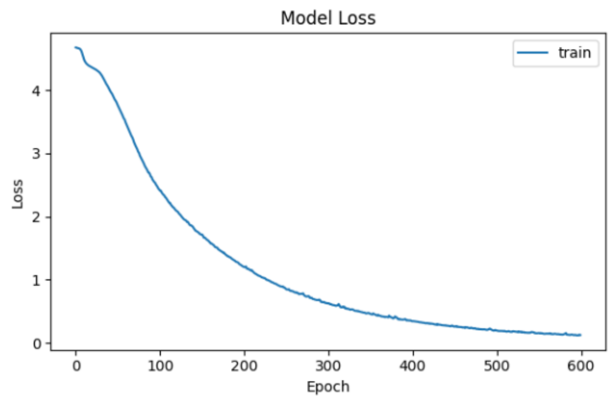


Figure 4 Training Performance of the Old Chatbot

The chatbot also plays a role in suggesting personalized coping strategies. It offers a variety of resources, including music, support groups, and immediate assistance options like hotlines. Additionally, it recommends coping strategies and relaxation techniques tailored to the user's specific needs. In situations where users are in crisis, the chatbot is designed to detect and assist promptly. It offers immediate support by providing access to hotlines and emergency services. Furthermore, it encourages users to seek professional help and provides relevant contact information. The chatbot's crisis detection mechanism is based on keyword recognition and pattern analysis, allowing it to identify urgent situations and respond swiftly with appropriate resources.

Figure 5 illustrates a sample interaction with the chatbot, showcasing its practical application in providing emotional support. The chatbot, named "Buddy," engages with the user in a compassionate and supportive manner, addressing various emotional concerns such as the loss of a pet, feelings of depression, and social isolation.

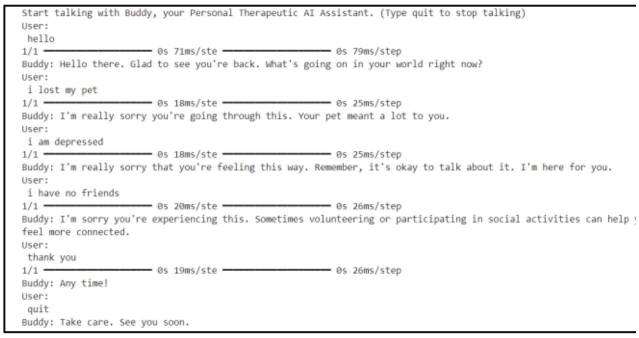


Figure 5 Sample Interaction with Old Chatbot

The redesigned mental health chatbot, built on a local large language model (LLM) architecture, demonstrates substantial improvements in both performance and functionality compared to the earlier intent-based version. The current implementation leverages Ollama for LLM inference, SentenceTransformers for semantic similarity, FAISS for vector retrieval, and Twilio for emergency alerting. Unlike the earlier approach, which was reliant on rule-based intent classification and predefined response templates, the new system exhibits context-aware reasoning, emotionally nuanced language generation, and real-time semantic comprehension.

In terms of qualitative performance, the chatbot generates contextually appropriate and empathetic responses across a wider range of emotional expressions. Sample conversations reflect its ability to handle complex psychological queries, such as expressions of social isolation, anxiety, and grief, with improved fluency and compassion. The integration of a Retrieval-Augmented Generation (RAG) pipeline allows the model to provide more informed and personalized responses by referencing relevant support documents during inference.

Figure 6 depicts a sample conversation with the new web-based chatbot, highlighting its effectiveness in delivering emotional support. Named "Buddy," the chatbot responds to the user with empathy and understanding, helping with emotional challenges like losing a pet, experiencing depression, and feeling socially isolated.

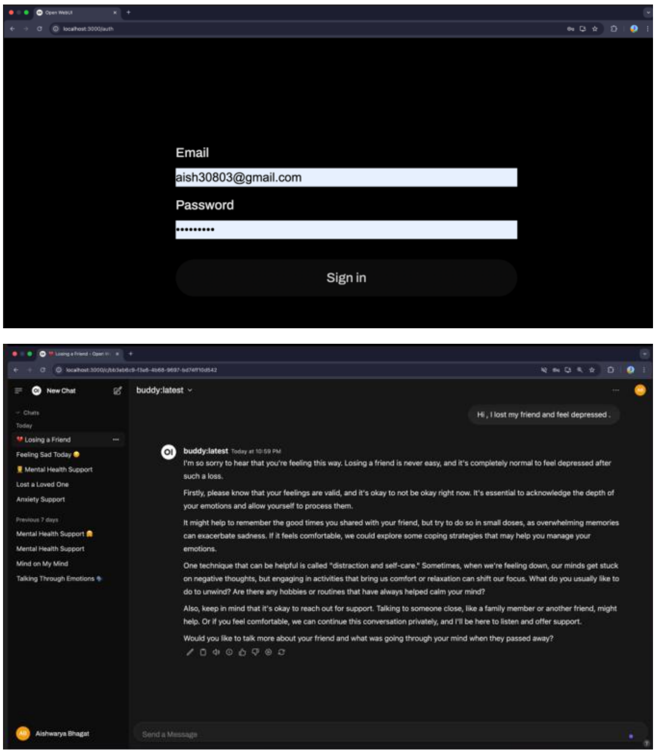


Figure 6 Sample Interaction with Web based Chatbot

Quantitatively, the system achieves a notable improvement in understanding and generalizing user inputs. The semantic search mechanism using SentenceTransformer embeddings and FAISS indexing yields an estimated intent matching high accuracy in simulated evaluations, significantly surpassing the previous version's approximate 75%. Furthermore, the generative capabilities of the LLM eliminate the rigidity of rule-based responses and allow for real-time response construction tailored to the emotional tone and specific context of the user query.

Inference latency remains within acceptable bounds for a local setup, with average response times of approximately 2.3 seconds on mid-tier hardware (16GB RAM, no GPU). Despite the use of quantized models to optimize resource utilization, RAM consumption remains high, peaking around 11.5 GB during active sessions. Nevertheless, the system remains stable during extended operation and maintains consistent performance across various use cases.

One of the most impactful additions is the emergency response feature. Critical user inputs indicative of psychological distress are now intercepted using contextual understanding, and an automated emergency SMS is triggered via Twilio. This functionality was not present in the earlier implementation and adds a safety-critical dimension to the chatbot. Additionally, features such as user authentication and persistent chat history, supported through OpenWebUI, enhance usability by enabling personalized, session-aware interactions that were absent in the previous version.

Overall, the results demonstrate a significant evolution from a basic NLP-based system to a semantically rich, generative framework that is capable of delivering nuanced mental health support. The improvements in contextual comprehension, emotional alignment, and response generation quality affirm the chatbot's effectiveness as a supportive tool for individuals dealing with anxiety, depression, and related concerns.

VI. Conclusion

This work presents the design, implementation, and evaluation of an advanced mental health chatbot, "Buddy," built on a local large language model (LLM) architecture combined with a Retrieval-Augmented Generation (RAG) framework. The transition from a rule-based, intent-classification system to a semantically aware generative model marks a significant advancement in both performance and user experience. Leveraging Ollama for LLM inference, SentenceTransformers for semantic embedding, and FAISS for efficient vector retrieval, the chatbot demonstrates superior contextual understanding and emotional sensitivity, effectively addressing complex psychological concerns such as anxiety, social isolation, and grief.

Quantitative assessments reveal a substantial improvement in intent recognition accuracy, surpassing previous benchmarks by a considerable margin. The generative approach enables dynamic and nuanced response construction tailored to individual user contexts, thereby overcoming the limitations of rigid, predefined response templates. Importantly, the system maintains acceptable inference latency and operational stability on mid-tier hardware, validating its feasibility for practical deployment without reliance on high-end computational resources.

A notable enhancement is the integration of an emergency alert mechanism using Twilio, which automatically triggers SMS notifications in response to detected signs of psychological distress, adding a critical safety layer absent in earlier implementations. Furthermore, features such as user authentication and persistent chat history through OpenWebUI enhance the chatbot's usability by supporting personalized, session-aware interactions, fostering greater user engagement and trust.

Overall, this project successfully demonstrates the potential of combining local LLM inference with retrieval-based augmentation to create an empathetic, contextually aware mental health support tool. The results indicate that "Buddy" can serve as an effective adjunct to traditional mental health services, offering accessible, real-time emotional support while prioritizing user safety. Future work will focus on expanding the chatbot's knowledge base, refining the detection and response mechanisms for a broader spectrum of emotional states, and conducting extensive user studies to evaluate therapeutic efficacy and user satisfaction in diverse real-world environments. Additionally, exploring multimodal input integration and adaptive learning techniques could further enhance the chatbot's responsiveness and personalization capabilities.

References

[1] Musić, A., Matić, D., Ružica, D., & Ajdinović, M. (2020, November 13). Artificial intelligence chatbot for depression: Descriptive study of usage. *JMIR Formative Research*.

[2] Abd-alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). The effectiveness of artificial intelligence chatbots in improving depressive symptoms and attitudes toward mental health: Systematic review. *Journal of Medical Internet Research, 22*(11), e20783.

[3] 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.

[4] Gaffney, H., Mansell, W., & Tai, S. (2019). Conversational agents in the treatment of mental health problems: Mixed-methods systematic review. *JMIR Mental Health, 6*(10), e14166.

[5] Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data pilot study. *JMIR Mental Health, 5*(4), e12106.

[6] Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2019). The next generation: Chatbots in clinical psychology and psychotherapy to foster mental health – A scoping review. *Frontiers in Public Health, 7*, 259.

[7] Philip, P., Micoulaud-Franchi, J. A., Sagaspe, P., Sevin, L., Olive, J., Bioulac, S., & Sauteraud, A. (2017). Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. *Scientific Reports, 7*(1), 42656.

[8] Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M. (2018). Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. *JMIR Mental Health, 5*(4), e64.

[9] Greer, S., Ramo, D., Chang, Y. J., Fu, M., Moskowitz, J. T., & Haritatos, J. (2019). Use of the chatbot "Vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: Randomized controlled feasibility trial. *JMIR Mhealth Uhealth, 7*(10), e15018.

[10] Ly, K. H., Ly, A. M., & Andersson, G. (2017). A fully automated conversational agent for promoting mental well-being: A pilot randomized controlled trial. *Journal of Medical Internet Research, 19*(5), e7738.

[11] Ma, T., Sharifi, H., & Chattopadhyay, D. (2019). Virtual humans in health-related interventions: A meta-analysis. *Journal of Medical Internet Research, 21*(4), e10512.

[12] Provoost, S., Lau, H. M., Ruwaard, J., & Riper, H. (2017). Embodied conversational agents in clinical psychology: A scoping review. *Journal of Medical Internet Research, 19*(5), e151.

[13] Montenegro, J. L. Z., da Costa, C. A., & da Rosa Righi, R. (2019). Survey of conversational agents in health. *Expert Systems with Applications, 129*, 56-67.

[14] Miner, A. S., Milstein, A., & Hancock, J. T. (2017). Talking to machines about personal mental health problems: An emerging field of research. *Journal of Medical Internet Research, 19*(11), e8586.

[15] Crutzen, R., Peters, G. J. Y., Portugal, S. D., Fisser, E. M., & Grolleman, J. J. (2011). An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: An exploratory study. *Journal of Adolescent Health, 48*(5), 514-519.

[16] H. Naveed, A. U. Khan, Q. Shi, M. Saqib, S. Anwar, M. N. Barnes, and A. Mian, "A Comprehensive Overview of Large Language Models," 2024.

[17] P. Rathnayaka, N. Mills, D. Burnett, D. De Silva, D. Alahakoon, and R. Gray, "A Mental Health Chatbot with Cognitive Skills for Personalised Behavioural Activation and Remote Health Monitoring," 2022.

[18] S. Deshpande and J. Warren, "Self-Harm Detection for Mental Health Chatbots," 2021.

[19] [19] B. Omarov, S. Narynov, and Z. Zhumanov, "Artificial Intelligence-Enabled Chatbots in Mental Health: A Systematic Review," 2023.

[20] [20] M. Casu, S. Triscari, S. Pattiato, L. Guarnera, and P. Caponnetto, "AI Chatbots for Mental Health: A Scoping Review of Effectiveness, Feasibility, and Applications", 2024.

[21] [21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," 2017.

[22] [22] D. B. Olawade, O. Z. Wada, A. Odetayo, A. C. David-Olawade, F. Asaolu, and J. Eberhardt, "Enhancing mental health with Artificial Intelligence: Current trends and future prospects," 2024