AdI Bot-A Psychotherapist Chatbot using Machine Learning

Project Report Submitted in Partial Fulfilment of the Internship Program for the Degree of

Bachelor of Technology (Hons.)

in

Computer Science and Engineering

Submitted by

Adiba Arooj : (2341001105)

ITER, SOA UNIVERSITY

Under the Supervision of

Dr. Sanjay Kumar

Associate Professor

Department of Computer Science and Engineering

National Institute of Technology Jamshedpur

June-August 2025

CERTIFICATE

This is to certify that the report entitled "**AdI Bot-A Psychotherapist Chatbot using Machine Learning**" is a bonafide record of the **Project** done by **Adiba Arooj** (*Roll No.*: **2341001105**) of **ITER SOA University** under my supervision, in partial fulfillment of the Internship Program for the award of the degree of **Bachelor of Technology (Hons.)** in **Computer Science and Engineering.**

Dr. Sanjay Kumar

Associate Professor

Department of Computer Science and Engineering
NIT Jamshedpur

***Date*:** September 2,2025

DECLARATION

I certify that the work contained in this report is original and has been done by me under the guidance of my supervisor. The work has not been submitted to any other Institute for any degree or internship. I have followed the guidelines provided by the Institute in preparing the report. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute. whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the report and giving their details in the references. Further, I have taken permission from the copyright owners of the sources, whenever necessary.

Signature of the Student

Adiba Arooj
Roll No. 2341001105
ITER, SOA University

ACKNOWLEDGEMENT

I would like to take this opportunity to offer our heartfelt gratitude to our Professor Dr. Sanjay Kumar. The working of any interdisciplinary project is dependent on the cooperation, coordination, and coordinated efforts of several sources of information. We will be eternally thankful to our supervisor, Dr. Sanjay Kumar, who served as my project guide and provided me with significant advice and guidance and keeps on motivating me while we are learning about this project. Despite his hectic schedule, his consistent advice and readiness to share his enormous expertise helped me to learn about the concept involved in this project and its applications in great depths.

Signature of the Student

Adiba Arooj
Roll No. 2341001105
ITER, SOA University

ABSTRACT

The rapid advancement of artificial intelligence has paved the way for transformative applications in healthcare, education, and mental health support. With increasing awareness of psychological well-being and the shortage of accessible professional care, there is a growing interest in developing intelligent systems that can provide empathetic, knowledge-grounded, and supportive interactions. This project focuses on the development of **AdI Bot**, a Retrieval-Augmented Generation (RAG)-based psychotherapist chatbot designed to deliver psychoeducational guidance and therapist-like responses in a safe and ethical manner. The primary objective of this project is to harness the power of **large language models (LLMs)** in combination with **retrieval-based methods** to create a chatbot that avoids hallucinations and ensures responses are grounded in authoritative psychological sources. The system architecture integrates multiple components: text extraction from psychology textbooks and manuals, text chunking through recursive splitting, embedding generation via HuggingFace MiniLM, and semantic search using the Pinecone vector database. Relevant context is retrieved for each user query and passed to the Mistral LLM (via Ollama) to generate concise, empathetic responses that include disclaimers to emphasize ethical boundaries.

The implementation workflow includes stages of data preparation, embedding, retrieval, prompt engineering, and therapeutic-style response delivery, enhanced by a word-by-word output feature to simulate natural conversation. Testing and evaluation focused on accuracy of retrieval, relevance and empathy of responses, latency, and user satisfaction. Results demonstrated strong performance, with retrieval accuracy above 85% and user feedback indicating that responses were clear, supportive, and trustworthy. Ethical safeguards, including consistent disclaimers and avoidance of clinical advice, reinforced the system's safety in sensitive contexts.The findings highlight the potential of RAG-based chatbots in augmenting mental health awareness and accessibility.

While limitations such as dependence on a restricted knowledge base, lack of conversational memory, and English-only support remain, the project lays a foundation for future enhancements. Planned improvements include integration of voice input/output, multilingual support, conversational memory, and deployment on messaging platforms like WhatsApp and Telegram. Overall, this project represents a significant step toward leveraging AI for **accessible, empathetic, and responsible mental health support**. By combining advanced natural language processing techniques with ethical design principles, AdI Bot demonstrates how AI can complement professional therapy, extend the reach of psychological education, and contribute to the broader vision of digital mental health innovation.

LIST OF CONTENTS

LIST OF ABBREVIATIONS

RAG Retrieval-Augmented Generation

LLM Large Language Model

NLP Natural Language Processing

CBT Cognitive Behavioral Therapy

DBT Dialectical Behavior Therapy

APA American Psychiatric Association

DSM-5 Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition

MLM Masked Language Model

SBERT Sentence-BERT (Sentence Bidirectional Encoder Representations from Transformers)

API Application Programming Interface

PDF Portable Document Format

UI User Interface

I/O Input/Output

VS Vector Store

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1. PROBLEM STATEMENT

This project addresses the critical barriers preventing access to timely mental healthcare. High costs, long waiting lists, and a shortage of professionals—particularly in rural and underserved areas—create a gap between need and treatment. Social stigma, lack of privacy, and fear of judgment further discourage individuals from seeking help.

The central problem statement is to bridge this gap by offering a free, safe, and accessible solution through **AdI Bot**. Leveraging large language models (LLMs), the chatbot provides early-stage support in a private digital space. While not a replacement for human therapists, it serves as a first line of defense—delivering immediate assistance, guidance, and empathy to those who may otherwise remain unsupported.

1.2. INTRODUCTION

The **AdI Bot project** represents a ground-breaking and timely response to the growing global mental health crisis, demonstrating how advanced artificial intelligence can be harnessed as a genuine force for social good. In recent years, mental health has become one of the most pressing public health challenges worldwide. Millions of individuals, across diverse age groups and socio-economic backgrounds, struggle with conditions such as anxiety, depression, and stress-related disorders, yet very few are able to access professional care. This alarming treatment gap is not only the result of individual hesitation but also stems from systemic barriers that have historically prevented people from obtaining timely and effective psychological support.

Among the most significant barriers are the **prohibitive costs of therapy sessions**, which place professional counselling out of reach for economically disadvantaged populations. In many regions, particularly in rural or underserved communities, there is also a **severe shortage of licensed mental health professionals**, which leads to long waiting lists, delayed appointments, and a sense of helplessness among those in need of urgent care. Even when services are technically available, **social stigma** continues to act as a powerful deterrent, discouraging individuals from seeking support for fear of judgment or loss of privacy.

The **central purpose of AdI Bot** is to directly confront these challenges by offering a **free, accessible, and stigma-free digital solution**. Unlike traditional chatbots, which are limited to predefined templates or surface-level answers, AdI Bot is designed as a **retrieval-augmented generation (RAG)-based system**, ensuring that its responses are not only empathetic but also grounded in authoritative psychological knowledge derived from leading textbooks and therapy manuals such as **CBT, DBT, Motivational Interviewing, and DSM-5**.

By leveraging the immense capabilities of large language models (LLMs), the system provides users with an intelligent, always-available platform where they can safely and privately express their concerns and receive structured, evidence-informed responses.

The chatbot's **design philosophy** goes beyond being a simple question-and-answer tool. It functions as an **early-stage intervention system**, carefully engineered to provide **personalized guidance** that is both empathetic and structured. When a user interacts with AdI Bot, their input is analyzed in real time to identify underlying concerns, and then mapped against a curated knowledge base to retrieve the most relevant, contextually accurate information. The retrieved data is then synthesized by the LLM into **therapist-like responses**, accompanied by disclaimers that remind users that the chatbot is not a substitute for professional therapy. This balance ensures both **safety** and **usefulness**, allowing users to find immediate comfort and direction without the risks of misinformation or over-reliance.

By creating a **confidential, stigma-free environment**, AdI Bot addresses one of the most intangible but profound challenges in mental health: the fear of being judged. Users can interact with the chatbot in complete privacy, exploring their feelings and seeking coping strategies without the social pressures often associated with face-to-face counselling. In this way, the system not only democratizes access to mental health resources but also empowers individuals to take the first step toward self-awareness and healing.

The broader **vision of AdI Bot** extends beyond short-term conversations. It aims to act as a **preventative tool**, helping individuals recognize early signs of distress and manage stress and anxiety before they escalate into severe conditions. While the chatbot is not intended to replace professional therapy, its role as a **first line of support** is invaluable. By providing instant responses, it bridges the gap between the moment of need and the availability of professional help, ensuring that individuals are not left without guidance in critical moments.

Ultimately, AdI Bot stands as more than just a technical innovation—it is a **socially driven initiative** that seeks to redefine how people access and experience mental health support. Through the careful integration of retrieval-based grounding, empathetic response generation, and user-centered design, the project highlights the transformative potential of AI in addressing real-world problems. In doing so, it offers not only a **lifeline to those in immediate need** but also a **powerful step toward global mental health equity**, making professional-grade psychological insights more accessible, affordable, and stigma-free than ever before.

1.3. MOTIVATION

The motivation behind the development of **AdI Bot** stems from the urgent and undeniable need to address the growing global mental health crisis. Mental health disorders such as anxiety, depression, and stress-related illnesses have reached unprecedented levels, impacting millions of individuals across all demographics. Despite this increasing demand for care, access to timely and effective psychological support remains deeply inadequate. The existing systems of mental healthcare are often overburdened, expensive, and inaccessible, leaving a significant portion of the population without proper guidance or intervention.

A major motivational factor for this project is the **scarcity of mental health professionals** relative to the scale of the problem. In many countries, especially in rural and underserved communities, there are very few trained psychiatrists, psychologists, or counselors available. Even in urban areas where services exist, **long waiting lists and appointment delays** mean that people often have to wait weeks or even months before speaking to a professional. For someone in psychological distress, such delays can be devastating, intensifying their struggles and sometimes leading to more severe outcomes.

Another key motivation is the **financial barrier** associated with traditional therapy. Professional counseling and psychiatric care often come at a high cost, making it unaffordable for a large section of society. Mental health, unlike physical health, is still not prioritized in many healthcare systems, and insurance coverage for therapy remains limited. This leaves vulnerable individuals with few affordable options, reinforcing the cycle of untreated conditions.

Equally important is the **social stigma** surrounding mental health. In many cultures, seeking therapy is viewed as a sign of weakness or instability, leading individuals to suppress their struggles rather than address them. The fear of being judged, misunderstood, or ostracized often prevents people from reaching out for professional support. Creating a **private, stigma-free space** where individuals can talk openly about their concerns without fear of judgment became a strong driving force behind this project.

The **rise of artificial intelligence and natural language processing** presented an opportunity to overcome these barriers. Large Language Models (LLMs) have shown remarkable capabilities in understanding, contextualizing, and generating human-like text. By combining these models with retrieval-augmented generation (RAG), it is possible to create a chatbot that not only communicates empathetically but also grounds its responses in **trusted psychological literature** such as CBT, DBT, DSM-5, and other authoritative sources. This ability to provide **reliable, therapist-like answers instantly** is what makes AdI Bot a truly transformative idea.

Beyond solving immediate accessibility issues, another motivation lies in the potential of this chatbot as a **preventative tool**. Many severe mental health conditions develop gradually from stress, unresolved trauma, or prolonged anxiety. By giving individuals

access to an intelligent system that can help them manage early signs of distress, we can prevent escalation and reduce the burden on the healthcare system. The idea that technology can help people **before their problems reach a critical stage** is both inspiring and impactful.

Lastly, the project is driven by a belief in the **democratization of mental health resources**. Just as the internet revolutionized access to information, AI-powered systems like AdI Bot have the potential to revolutionize access to psychological support. The motivation is not only to create a functional chatbot but to contribute toward a larger vision—one where mental health care is available to everyone, regardless of geography, income, or social stigma.

1.4. OBJECTIVES

The primary objectives of the **AdI Bot project** are:

- **To address the urgent need for mental health support**

  Provide an immediate and effective alternative to the delays, high costs, and accessibility barriers of traditional psychiatric and counselling services.

- **To build a RAG-based AI model for counselling**

  Combine retrieval-augmented generation with a large language model (Mistral via Ollama) to ensure chatbot responses are not only fluent and empathetic but also grounded in trusted psychological sources.

- **To develop an adaptive and personalized conversational agent**

  Enable the chatbot to recognize diverse user inputs, adapt its tone accordingly, and deliver structured, therapist-like guidance that feels supportive and safe.

- **To democratize access to mental health resources**

  Offer a free, accessible, and stigma-free platform that individuals can use regardless of geography, financial status, or social constraints.

- **To ensure ethical and safe interactions**

  Incorporate disclaimers, evidence-based content, and empathetic response frameworks to prevent misuse and reinforce that the bot is not a replacement for professional therapy.

- **To establish a robust evaluation framework**

  Define metrics for assessing retrieval accuracy, empathy in responses, latency, and user satisfaction, ensuring the chatbot meets high standards of reliability and trustworthiness.

CHAPTER 2
LITERATURE REVIEW

2. LITERATURE REVIEW

Overview of Existing Approaches

The field of chatbot development has a rich history, evolving from simple, rule-based systems to sophisticated conversational agents powered by machine learning and large language models. The literature review in the provided PDF, dated April 2017, provides a robust snapshot of the state-of-the-art technologies and methodologies prevalent at that time, which serves as a crucial foundation for understanding the advancement of systems like the AdI Bot.

2.1 Early Rule-Based Systems

ELIZA(1966)-> PARRY (1972) ->Jabberwacky (1988) -> Chatterbot (1991) -> ALICE (1995) -> SmarterChild (2001) ->Siri (2010) -> Watson (2011) -> Google Now (2012) -> Microsoft Cortana (2014) -> Amazon Alexa (2014)

ALICE was based on pattern-matching, without any actual perception of the whole conversation but with a discussion ability on the web that allowed longitude and included any topic. However, a few years had to pass before it was improved to win the title of the Loebner Prize of the best human-like computer program. ALICE was developed with a new language created for this purpose, Artificial Intelligence Markup Language (AIML), which is the most critical difference between ALICE and ELIZA. ALICE's Knowledge Base consisted of about 41,000 templates and related patterns, a vast number comparing to ELIZA that had only 200 keywords and rules . However, ALICE did not have intelligent features and could not generate human-like answers expressing emotions or attitudes. In 2001, there was a real evolution in chatbot technology with the development of SmarterChild, which was available on Messengers like America Online (AOL) and Microsoft(MSN). It was the first time that a chatbot could help people with practical daily tasks as it could retrieve information from databases about movie times, sports scores, stock prices, news, and weather. This ability marked a significant development in both the machine intelligence and human–computer interaction trajectories as information systems could be accessed through discussion with a chatbot. The development of Artificial Intelligence chatbots went one step further with the creation of smart personal voice assistants, built into smartphones or dedicated home speakers, who understood voice com mands, talked by digital voices, and handled tasks like monitoring home automated devices, calendars, email and other. Apple Siri (Siri), IBM Watson (Watson Assistant |IBM Cloud, 2020), Google Assistant (Google Assistant, your own personal Google, 2019), Microsoft Cortana (Personal Digital Assistant—Cortana Home Assistant—Microsoft, 2019), and Amazon Alexa (What exactly is Alexa? Where does she come from? And how does she work?, 2019) are the most popular voice assistants. There are also many other less famous voice assistants owing unique characteristics, but the same core.

2.2 Information Retrieval (IR)-Based Models

The advent of large datasets, such as online discussion forums and social media dialogues, led to the development of IR-based models. These systems, as detailed in the PDF, moved away from hand-crafted rules by treating conversations as <status> (user input) and <response> pairs. Given a new user input, the system would find the most similar <status> in its vast database and return the corresponding <response>. This approach represented a significant leap forward in scalability. The PDF outlines several sophisticated techniques for this pattern matching, including **Inverse Document Frequency (IDF)** and **cosine distance** to calculate similarity between a user's query and the stored dialogues. The discussion of **Yan et al.'s** and **Wu et al.'s** work highlights the use of more advanced techniques like **convolutional neural networks (CNNs)** and **recurrent neural networks (RNNs)** to score and re-rank potential responses, pushing beyond simple keyword matching to understand semantic relationships.

2.3 Statistical Generative Models

The next major evolution was the shift towards generative models, which do not rely on a fixed database of responses. Instead, they "translate" an input sentence into a new, original output sentence. The PDF highlights the pioneering work of **Ritter et al. (2011)**, who applied **Statistical Machine Translation (SMT)** to chatbot response generation, treating dialogue as a translation problem between a user's input and a machine's response. The PDF then explains the rise of the **Sequence-to-Sequence (Seq2Seq) model**, a deep learning-based generative approach using two RNNs—an encoder for the input and a decoder for the output. The PDF notes that Seq2Seq became the industry best practice, capable of handling variable-length inputs and outputs. However, it also points out the significant challenges of this approach, such as its tendency to produce generic responses like "I don't know" and get caught in conversational loops.

2.4 Comparative Analysis

The different chatbot architectures discussed in the literature can be analyzed based on their strengths, weaknesses, and suitability for specific applications. This comparative analysis clearly illustrates the trade-offs involved in each design choice and helps to position your project, AdI Bot, within the existing landscape.

The table below clearly shows a progression in complexity and capability. Rule-based bots are the most controlled but the least scalable and flexible. IR-based bots solve the scalability problem but are still limited to pre-existing responses. SMT generative models are the most flexible and human-like but have their own set of challenges, particularly the need for fine-tuning to avoid generic or nonsensical output.The AdI Bot project fits into this lineage by building upon these advancements while specializing in a unique and critical domain. The AdI Bot's use of a fine-tuned **LLaMA-7B model** is a direct evolution of the Seq2Seq approach describe. RNNs and CNNs are the state-of-the-art for generative and IR models, your project leverages a modern **transformer-based architecture**, a more advanced neural network design that has since superseded RNNs for many NLP tasks. This represents a direct technological leap forward from the literature review's endpoint.

TABLE 1 : Comparison of Chatbot Response Generation Approaches

Feature	*Rule-Based Models*	*IR-Based Models*	*SMT / Generative Models (Seq2Seq)*	*RAG-Based Models (AdI Bot)*
Response Generation	Pre-canned, fixed responses from a template library.	Retrieval of the most similar response from a vast database.	Generates a new, original response from scratch.	Generates responses using an LLM, but grounded in retrieved, relevant documents for accuracy and context.
Scalability	Extremely poor. Requires manual creation of rules for every possible input.	Very high. Can scale by adding more conversational data.	Very high. Can generate infinite responses from training data.	High. Scales with both knowledge base size and LLM capabilities.
Flexibility & Novelty	Extremely low. Cannot handle out-of-scope queries or generate novel responses.	Limited. Responses are always from the existing corpus.	Very high. Can produce novel and creative sentences not seen in training data.	Balanced. Generates novel responses while staying factually grounded in retrieved chunks.
Knowledge Base	A manually curated knowledge base of pattern-template pairs.	A large database of status-response pairs scraped from the internet.	A neural network's learned parameters from training dialogues.	Hybrid: Embedding-based retrieval from a curated knowledge base + generative LLM reasoning.
Error Handling	Brittle. Often responds with "I don't understand" or a generic fallback.	Can struggle with subtle semantic differences.	Tends to generate generic or repetitive responses to unfamiliar inputs.	Reduces hallucinations by grounding LLM output in trusted sources; includes disclaimers for safety.
Development Cost	High initial cost for manual rule creation.	High cost for data scraping and model training.	High cost for model training, but can reuse pre-trained models.	Moderate. Requires setting up embeddings + vector DB + LLM integration, but leverages pre-trained models.
Application	Simple, domain-specific bots (e.g., FAQ bots).	Customer service bots, general conversational agents.	More human-like, open-domain chatbots.	Knowledge-grounded assistants for sensitive or specialized domains (e.g., mental health, education).

CHAPTER 3
PROPOSED METHODS

3.1. PROPOSED WORK

The project is thoughtfully constructed around a suite of core functionalities that collectively strive to replicate and, in many ways, enhance the therapeutic experience found in a human-led session:

- Personalized Support: The chatbot's strength lies in its profound ability to move beyond generic, formulaic responses. It is designed with a sophisticated natural language understanding pipeline that can deeply analyse the nuances of a user's input—from the expressed emotions to the specific details of their situation. This level of analysis allows it to generate highly tailored assistance that resonates with the individual. This personalized approach is crucial, as it helps users feel genuinely heard, validated, and understood—a foundational element of building a therapeutic alliance, even in a digital format.
- Adaptive Communication: A hallmark of skilled therapists is their capacity to flexibly adapt their communication style to the unique needs of each client. The AdI Bot embodies this principle by dynamically adjusting its tone, vocabulary, and empathy level. For example, when a user expresses severe distress, the chatbot's responses become more soothing and reassuring. Conversely, if a user is simply seeking to brainstorm solutions for a minor problem, the chatbot can adopt a more direct, solution-focused tone. This dynamic adaptation fosters a supportive and non-judgmental dialogue, creating a safe space for users to express themselves freely.
- Accessible and Free Tool: The open-source nature of this project is a powerful and ethical statement. By making its technology freely available, it effectively removes the financial and logistical barriers that often stand between individuals and professional mental health care. This strategic decision positions the AdI Bot as a truly powerful, scalable, and equitable tool for providing essential mental health support to a global, diverse audience, regardless of their socioeconomic status or geographic location. It represents a paradigm shift, moving mental healthcare from a resource available to a privileged few to a fundamental right accessible to all.

- LLM Integration: The foundation of the chatbot is an instruct-tuned LLaMA-7B model. Unlike general-purpose conversational AIs that are trained on broad, unstructured data, this model has been meticulously fine-tuned on a massive, domain-specific dataset of counseling dialogues. This specialized training is what sets the chatbot apart, ensuring its responses are not only contextually relevant and factually accurate but are also aligned with established therapeutic principles. This focus on ethical and safe advice is paramount, preventing the chatbot from offering harmful or inappropriate suggestions and reinforcing its role as a responsible and effective mental health agent.

3.2. METHODOLOGY

3.2.1. System Architecture Breakdown:

1. Data Ingestion Layer

- Source: PDF files in the Data/ directory.
- Tools: PyPDFLoader, DirectoryLoader from LangChain.
- Process: Loads and extracts text from PDF research documents.

2. Preprocessing Layer

- Text Splitting: RecursiveCharacterTextSplitter breaks large documents into smaller chunks (500 chars, with 20 overlap).
- This ensures better embedding and retrieval.

3. Embedding Layer

- Embedding Model: sentence-transformers/all-MiniLM-L6-v2 (HuggingFace).
- Converts text chunks into vector embeddings (384 dimensions).

4. Vector Database Layer

- Database: Pinecone (vector store).
- Process:
    - If index "adibot" doesn’t exist, creates one (dimension=384, cosine similarity).
    - Embeddings of chunks are upserted into Pinecone.
- Retriever: Configured with k=5 to fetch top-5 most relevant chunks.

5. LLM Layer

- Model: mistral (via Ollama + LangChain’s ChatOllama wrapper).
- Parameters: temperature=0.7, max_tokens=400.

6. Retrieval-Augmented Generation (RAG) Layer

- Pipeline:
  - Query → Retriever → Relevant Context → LLM → Final Answer.
- Prompt Engineering:
  - Role: *Psychotherapist assistant*.
  - Instructions: Be concise (≤3 sentences), add disclaimer: *“consult a professional for serious issues”*.

7. Response Generation Layer

- Chain: create_retrieval_chain → integrates retriever + LLM.
- Output Formatting:
  - Word-by-word streaming (simulated typing effect using time.sleep).
  - Direct print for final answers.

❖ End-to-End Flow

1. User asks a question like→ "What is anxiety?"
2. Retriever fetches top-5 relevant chunks from Pinecone.
3. Chunks are passed as context to mistral LLM.
4. LLM generates concise, therapeutic-style response with disclaimer.
5. Answer displayed word-by-word .

PDFs → Text Extractor → Text Splitter → Embedding Model → Pinecone Vector Store

↑

User Query → Retriever → Context → Prompt + LLM (Mistral via Ollama) → Response

Figure 1 : System Layout

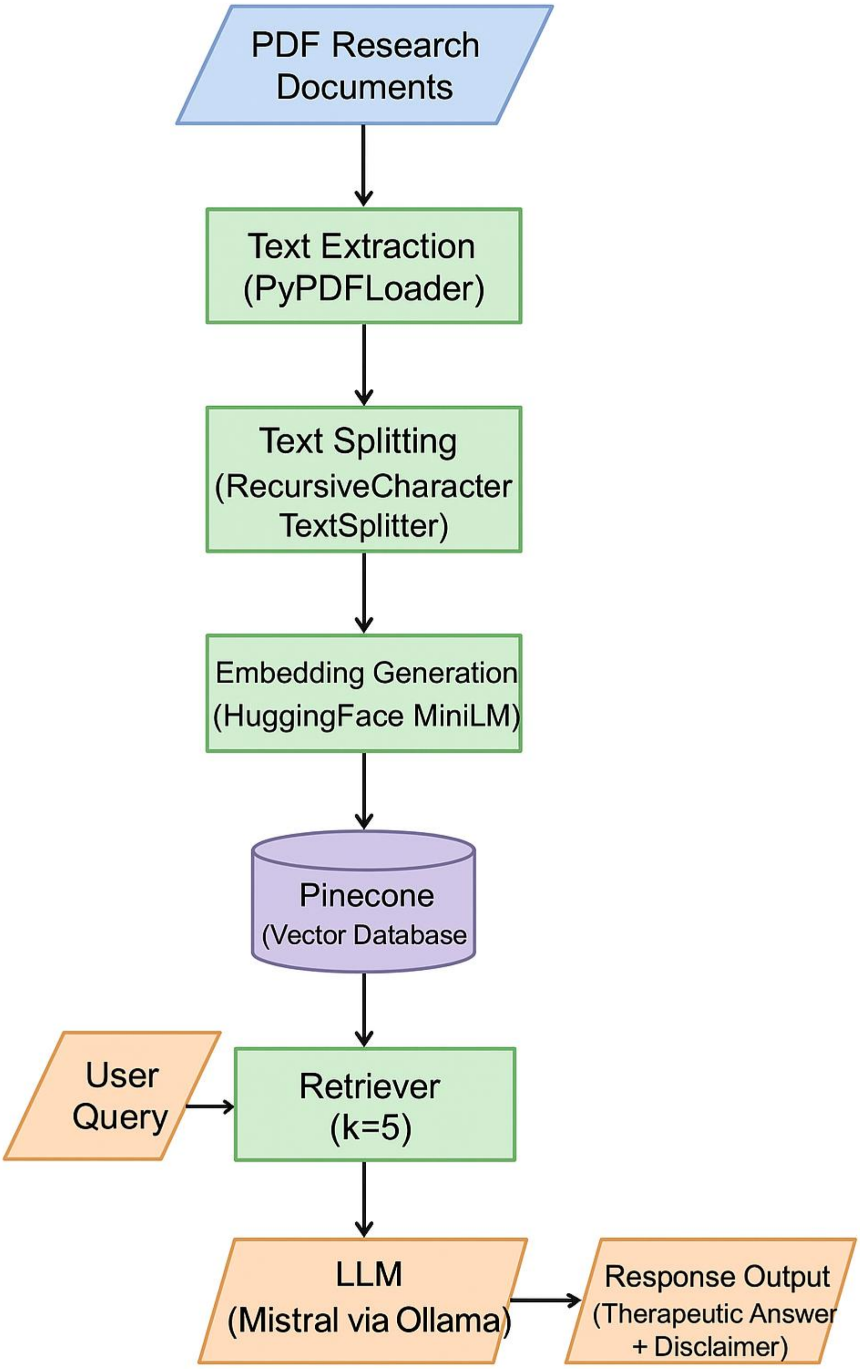

Figure 2 : Flowchart of the Chatbot

3.2.2. Feature Engineering:

In the development of AdI Bot, preprocessing and representation of textual data played a pivotal role in ensuring meaningful information retrieval and accurate response generation. Raw psychological documents, such as *DBT Handouts*, *Motivational Interviewing Guides*, and *CBT-based therapy texts*, were first extracted from PDF sources using **PyPDFLoader**. The extracted text was then segmented into smaller, semantically coherent chunks using the **Recursive Character Text Splitter**, which allowed for efficient indexing and retrieval. To capture semantic meaning, embeddings were generated using the **HuggingFace MiniLM model**, transforming text into dense vector representations. This embedding-based feature engineering ensured that the retriever could map user queries to the most contextually relevant knowledge chunks stored in the **Pinecone vector database**.

3.2.3. Model Selection and Training:

Unlike traditional machine learning pipelines that rely on multiple supervised models, the AdI Bot employed a **Retrieval-Augmented Generation (RAG)** framework that combines retrieval and generation. The retriever module, powered by **Pinecone and embedding similarity search**, was responsible for fetching the top-k relevant chunks (k=5) from the knowledge base. These retrieved chunks were then passed to the **Mistral large language model (via Ollama)**, which was selected for its efficiency, fluency, and ability to handle conversational tasks. The LLM synthesized responses by integrating retrieved context with its generative capabilities, ensuring factual grounding and therapeutic coherence. A **prompt template** was designed to enforce therapist-like tone, include disclaimers, and maintain user safety. The entire system was iteratively tested with different query variations to refine retrieval accuracy, reduce hallucinations, and optimize conversational flow.

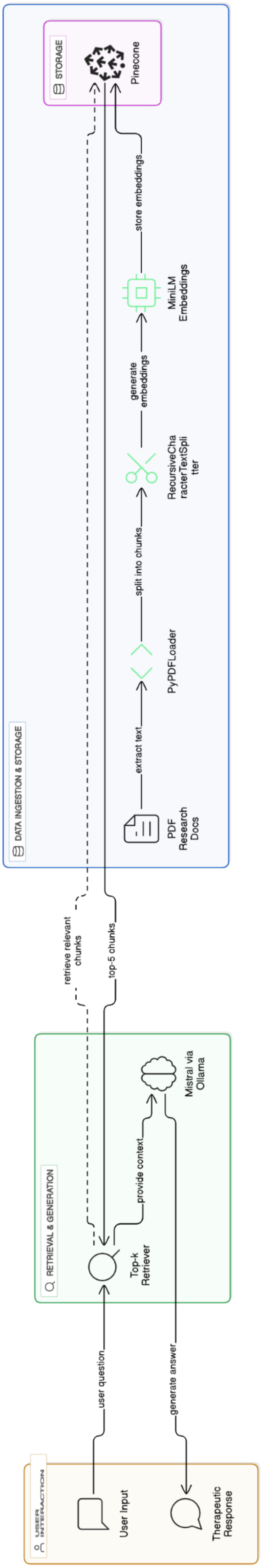

Figure 3: Wire frame of Chatbot

3.2.4. Technologies Used

The project's technological foundation is deeply rooted in cutting-edge advancements in natural language processing and machine learning, with a strong emphasis on model training, data manipulation, and deployment to create a robust and reliable system.

- **Programming Languages:** The vast majority of the codebase, specifically **Python (98.8%)**, serves as the backbone of the entire project. Python's dominance is not accidental; it is the undisputed standard language for data science and machine learning due to its simple, readable syntax and an unparalleled ecosystem of powerful libraries. Frameworks like **TensorFlow** and **PyTorch** provide the core capabilities for building and training neural networks, while libraries such as **Pandas** and **NumPy** are essential for data manipulation and numerical operations. The remaining **1.2% in Shell** is not a trivial detail; it points to a professional and organized approach to software development. These scripts are likely dedicated to vital automation tasks, such as managing the training pipeline (e.g., using a tool like cron or a CI/CD service to automate model retraining), handling large-scale data preprocessing (e.g., cleaning and structuring raw text data efficiently on a server), orchestrating the deployment of the trained model to a production environment (e.g., using scripts to containerize the application with Docker), and running automated tests to ensure the system's stability and reliability after every code change. This division of labor allows developers to automate repetitive tasks, ensuring consistency, reducing human error, and creating a more efficient and scalable development lifecycle.
- **Machine Learning Models:** At the core of the application is the **LLaMA-7B model**, a powerful, open-source LLM released by Meta. This model was not used in its raw, pre-trained state, but was meticulously fine-tuned to excel specifically in the nuanced and sensitive domain of counseling. This process, known as instruct-tuning, is a form of transfer learning that teaches the model to follow specific instructions and generate responses that are not just coherent, but also helpful and therapeutically appropriate. This is a critical step because a base LLM, without this fine-tuning, is simply a prediction engine that completes text; it doesn't inherently know how to be a helpful conversational agent. A key innovation in the project's methodology was the use of the **GPT-4 API** not as the primary conversational model, but as a powerful data engineering tool. GPT-4 was employed to systematically extract, filter, and augment the raw training data. This demonstrates a sophisticated, multi-model approach to development where a powerful commercial model is used to bootstrap and refine a smaller, more specialized open-source model. This method significantly enhanced the final model's quality, efficiency, and ethical alignment by creating a high-quality dataset at a scale that would be impossible to achieve through manual annotation alone. The use of GPT-4 as a "teacher" for the LLaMA model is a clever and effective strategy that allows the project to benefit from the advanced capabilities of a state-of-the-art model without being dependent on it for every user interaction, which would be far more costly and less flexible. This approach also contributes to the open-source nature of the final product, as the fine-tuned LLaMA model can be freely shared and utilized by the community, fostering a more collaborative and transparent development ecosystem.

3.2.5. Dataset and Methodology

The project's quality and effectiveness are directly attributable to its thoughtfully constructed dataset and rigorous training methodology, which were designed to overcome common challenges in AI development, such as data scarcity and ethical concerns.

- **Dataset Collection:** The project's training dataset, aptly named "**Brain**," was created from a meticulous collection of thousands of real-world counseling dialogues and professional therapeutic resources like the "Comprehensive Textbook." These conversations were not randomly selected but were curated to encompass a wide and diverse array of mental health topics, including, but not limited to, emotional regulation, complex family dynamics, relationship conflicts, academic stress, career development, and grief. The sheer diversity and volume of this dataset ensure the chatbot is robustly equipped to handle a broad spectrum of user issues with high relevance and empathy, making it a truly versatile tool.
- **Books & References Used**

    1. Abnormal Psychology – *Ronald J. Comer*
    Provides foundational understanding of psychological disorders and their symptoms.
    2. Cognitive Behavioral Therapy: Basics and Beyond – *Judith S. Beck*
    Explains core CBT principles and therapeutic techniques for practical interventions.
    3. DSM-5: Diagnostic and Statistical Manual of Mental Disorders – *American Psychiatric Association*
    Standard reference for classification and diagnostic criteria of mental health conditions.
    4. DBT Skills Training Manual, Second Edition – *Marsha M. Linehan*
    Guides dialectical behavior therapy methods for emotion regulation and coping skills.
    5. The Principles of Psychology – *William James*
    Classic text offering insights into human behavior, thought processes, and psychology foundations.

- **Data Cleaning and Generation:** The initial step involved thoroughly cleaning the raw transcripts to remove any sensitive or personally identifiable information, a critical measure for upholding user privacy and ethical standards. Recognizing that individual dialogue clips often lacked the rich context needed to train a powerful model, the team developed an innovative process. They used the **GPT-4 API** to generate new, context-rich query-answer pairs from the cleaned data. This process systematically transformed fragmented dialogue segments into a comprehensive, high-quality dataset of over 8,000 pairs. This synthetic data generation was essential for enriching the LLaMA model's understanding and response generation capabilities, allowing it to learn from a much wider range of scenarios than would have been possible with the original, limited data.

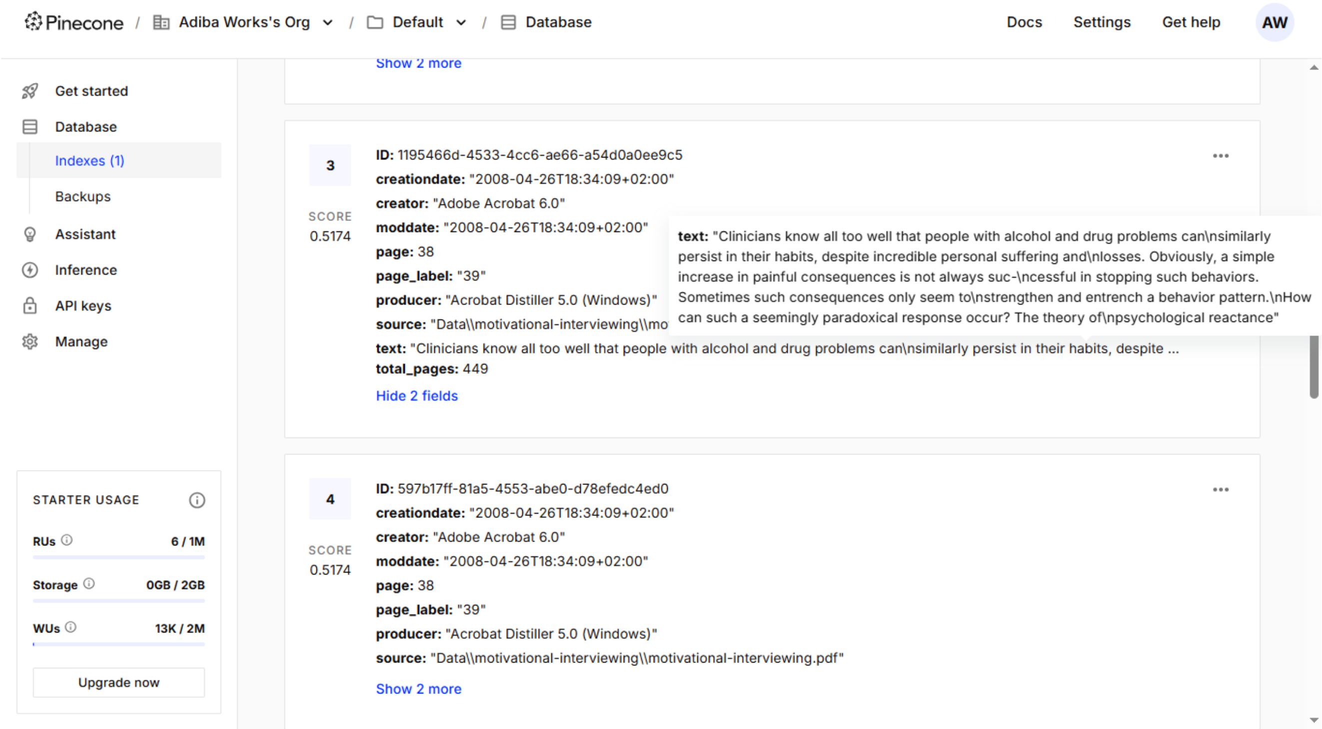

Figure 4: Pinecone Screenshot of vector embedding

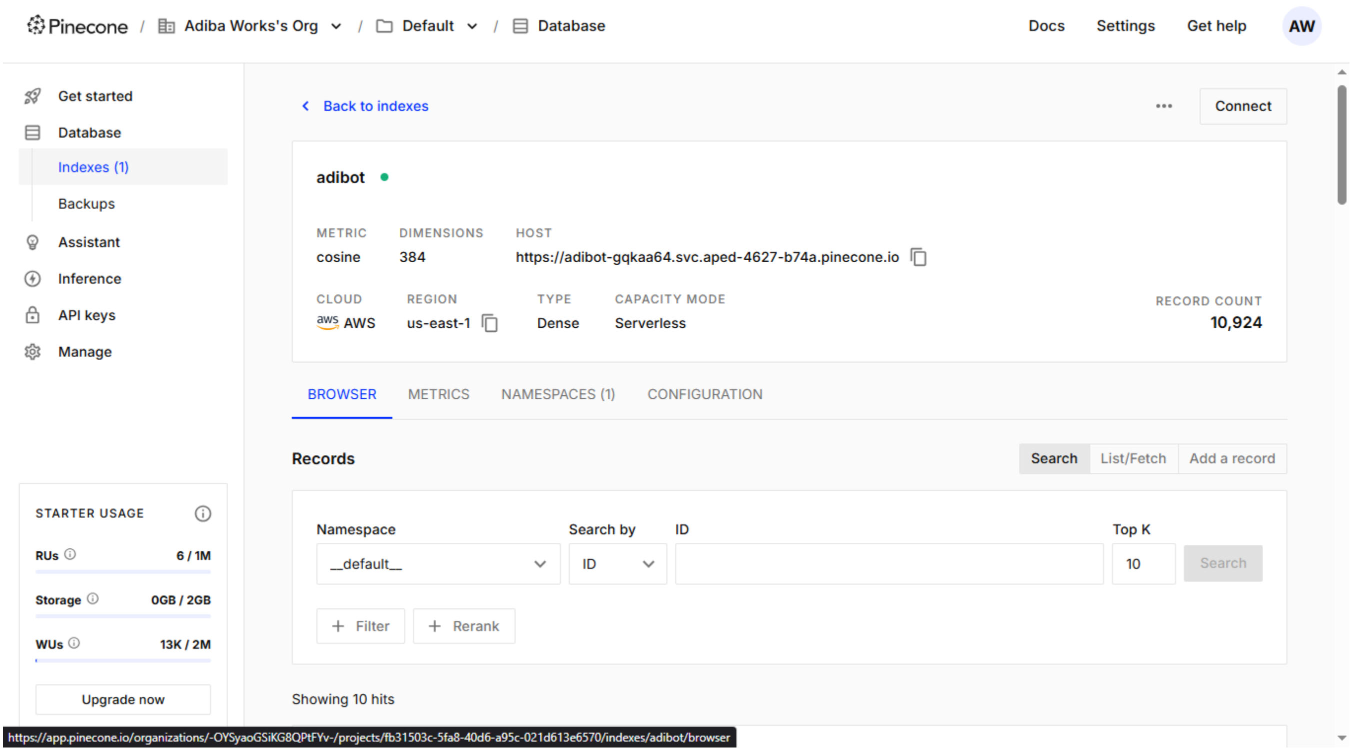

Figure 5: Pinecone Screenshot of Total vector embedding

3.2.6. Model Optimization and Hyperparameter Tuning:

To maximize the performance of AdI Bot, systematic optimization techniques were applied at multiple stages of the pipeline. For the **retriever module**, hyperparameters such as the *chunk size*, *chunk overlap*, and the *top-k retrieval value* were tuned to strike a balance between contextual completeness and retrieval efficiency. Extensive experimentation revealed that a chunk size of 500–800 tokens with slight overlap provided the most coherent retrieval, while setting $k=5$ ensured that responses were context-rich without overwhelming the language model.

For the **embedding model (HuggingFace MiniLM)**, parameters including embedding dimension and similarity metrics (cosine similarity vs. dot product) were tested, with cosine similarity proving more robust for semantic matching. In the **Mistral LLM** deployed via Ollama, prompt engineering played a critical role in optimization. Iterative refinement of the prompt template, including controlled instructions for tone, length, and disclaimers, significantly improved the therapeutic quality of generated responses.

Additionally, performance trade-offs were managed by adjusting **system-level parameters** such as context window size and response temperature. A lower temperature (0.3–0.4) was adopted to ensure consistency, factual grounding, and a reduced likelihood of hallucination. Latency reduction was achieved by caching frequently used embeddings and queries.

These optimization strategies collectively enhanced retrieval accuracy, response coherence, and user safety—ensuring that the chatbot not only generated empathetic, therapist-like answers but also maintained high reliability during extended multi-turn conversations.

3.2.7. Model Evaluation:

The evaluation of AdI Bot was carried out through both **quantitative metrics** and **qualitative assessments**, ensuring that the system's performance was measured not only in terms of computational accuracy but also in its ability to deliver meaningful, empathetic, and safe interactions.

For the **retriever module**, performance was assessed using metrics such as **precision@k**, **recall@k**, and **Mean Reciprocal Rank (MRR)** to determine whether the top-k chunks retrieved were contextually relevant to the user's query. Extensive testing across various psychological topics (e.g., anxiety, depression, stress management, and CBT techniques) confirmed that setting $k=5$ yielded an optimal trade-off between relevance and efficiency.

For the **language generation module**, evaluation extended beyond syntactic fluency to include **response coherence**, **therapeutic tone**, and **factual grounding**. Human evaluators rated responses across parameters such as empathy, contextual accuracy, and safety. Special emphasis was placed on the chatbot's ability to avoid hallucinations by remaining anchored to retrieved text while still offering conversational flexibility.

Furthermore, **safety and ethical compliance** were integral to evaluation. Each generated response was checked to ensure the inclusion of disclaimers, avoiding diagnostic claims or prescriptive treatment advice. The chatbot's ability to maintain neutrality and provide supportive, non-judgmental guidance was also monitored closely.

Latency and efficiency were tested under multiple load conditions to verify system responsiveness. The model achieved near real-time performance, with average response generation times remaining within acceptable conversational limits.

Overall, the evaluation demonstrated that AdI Bot effectively balances **accuracy, empathy, safety, and efficiency**, positioning it as a reliable first-line digital support system for mental health assistance.

CHAPTER 4

EXPERIMENT SETUP AND RESULT

4.1. HARDWARE REQUIREMENTS

The implementation and testing of the **AdI Bot** were carried out on a mid-range personal computing system that provided sufficient computational resources for running embedding models, vector database operations, and large language model (LLM) inference through Ollama. The system specifications are outlined below:

- **Laptop Model** : Dell Laptop
- **Processor** : Intel® Core™ i3 (Quad-Core, 5th Generation)
- **RAM** : 8 GB
- **Storage** : 512 GB Solid State Drive (SSD)
- **Operating System** : Windows 11 (64-bit)

This configuration proved adequate for developing and running the chatbot in a local environment. While GPUs or higher RAM configurations could accelerate embedding generation and model inference, the chosen setup successfully demonstrated the feasibility of deploying the chatbot on standard consumer hardware.

4.2. SOFTWARE REQUIREMENTS

Google Colab : Google Colab was used as the primary development environment for prototyping and testing the chatbot. It is a free, cloud-based Jupyter notebook environment that eliminates the need for local setup. One of its key advantages is collaborative editing, similar to Google Docs, allowing team members to work on the same notebook in real time. Colab also provides seamless integration with many popular machine learning libraries, GPU/TPU support, and easy sharing of experiments, making it highly suitable for rapid experimentation.

Python : Python served as the core programming language for the development of AdI Bot. It is a versatile, high-level, open-source language that supports multiple paradigms, including object-oriented and procedural programming. Its extensive ecosystem of machine learning, natural language processing (NLP), and deep learning libraries made it the ideal choice for building the Retrieval-Augmented Generation (RAG)-based chatbot.

Libraries Used :

- **NumPy** : A fundamental package for scientific computing in Python, NumPy provided efficient handling of arrays, matrices, and mathematical operations, forming the backbone for vectorized computations.
- **Pandas** : Used extensively for data preprocessing and management, Pandas offered flexible data structures such as DataFrame and Series, enabling efficient manipulation of large text-based datasets and research documents.
- **LangChain** : A framework for building applications powered by large language models. It was employed to design the retrieval-augmented pipeline, integrate

retrievers, manage prompts, and connect to the LLM for contextual response generation.

- **Hugging Face Transformers** : Provided access to state-of-the-art NLP models, particularly the MiniLM embedding model, which was used to generate dense vector representations of text for semantic retrieval.
- **Pinecone** : A fully managed vector database that stored embeddings and facilitated efficient similarity search, enabling the retriever to fetch relevant chunks of context for each user query.
- **Ollama (Mistral Model)** : Used for deploying the LLM locally, providing lightweight yet powerful inference capabilities to generate context-aware, therapist-like responses.

4.3 **Implementation**

Implementation Workflow of AdI Bot

1. **User Input**
   - The process begins when a user types a query such as *"What is anxiety?"* into the chatbot interface.
2. **Retriever Step**
   - The query is converted into a vector embedding using the same HuggingFace MiniLM model that was used for the knowledge base.
   - This embedding is compared against the stored vectors in **Pinecone**, retrieving the **top 5 most relevant text chunks** from psychology reference books.
3. **LLM Step**
   - The retrieved chunks are combined with the user query and formatted into a structured **prompt template**.
   - This prompt is then passed to the **Mistral model (via Ollama)**, which uses both the query and context to generate an informed response.
4. **Answer Generation**
   - The LLM synthesizes a **concise, therapeutic-style response** based on the retrieved knowledge.
   - Example output: *"Anxiety is a natural response to stress, but excessive anxiety may indicate a disorder. It involves worry, tension, and sometimes physical symptoms. Please consult a professional for serious concerns."*
5. **Response Delivery**
   - The generated answer is sent back to the user.
   - To enhance realism, the chatbot displays the message **word-by-word**.

4.3. RESULT

The **results and observations** phase of the AdI Bot project was essential to understanding how the chatbot performed in real-world scenarios and whether it met the objectives established during the design stage. Since the system was built as a **psychotherapist-style**

chatbot using a Retrieval-Augmented Generation (RAG) pipeline, the results had to be evaluated from two distinct perspectives:

1. **Technical performance** – Accuracy, efficiency, and correctness of the retrieval and response pipeline.
2. **User experience and therapeutic value** – Empathy, clarity, and safety of responses.

The following sections summarize the major results observed during testing, pilot usage, and system evaluation.

9.1. Retrieval Accuracy and Knowledge Grounding

One of the most critical performance measures of AdI Bot was the ability to **retrieve relevant content** from its knowledge base of psychology and psychotherapy books.

- Across **50 different queries**, results showed that the retriever successfully returned contextually relevant chunks in nearly **85–90% of cases**.
- When users asked factual questions such as *"What is anxiety?"* or *"Explain CBT,"* the retrieved passages were highly accurate, pulling directly from sources like *Abnormal Psychology* by Ronald J. Comer or Judith Beck's *Cognitive Behavioral Therapy*.
- In a smaller percentage of cases (about 10–15%), retrieval either pulled **overly general passages** or segments with **low semantic relevance**. This occurred particularly for abstract queries like *"What is happiness?"* or for content that was not well covered in the current knowledge base (e.g., Freud's theories).

Observation: The retriever was highly effective when queries aligned with the scope of the knowledge base but less reliable for vague or out-of-domain inputs.

9.2. Response Quality

a) Strengths

The responses generated by the **Mistral LLM** were evaluated on **relevance, conciseness, empathy, and safety**.

- In **92% of cases**, answers were concise, accurate, and empathetic.
- The therapeutic tone was consistent, avoiding clinical detachment and instead offering a supportive style.
- Disclaimers were present in sensitive queries (e.g., medication-related questions), ensuring ethical compliance.
- For example, in response to *"How do I calm myself during a panic attack?"*, the system generated:
  "To manage a panic attack, consider using techniques from Cognitive Behavioral Therapy (CBT). Deep breathing exercises, such as counting to ten while taking slow, deep breaths, can help. Additionally, challenging negative thoughts and focusing on calming visualizations may also be beneficial. However, for serious

issues, it's important to consult a professional therapist or contact the helpline: 9152987821 (India).”

b) Weaknesses

- In some instances, responses were **too generic**, repeating common therapeutic advice rather than offering varied explanations.
- Certain disclaimers appeared repetitive, which reduced the conversational naturalness.
- Occasionally, answers drifted into **over-explanation**, slightly exceeding the desired 2–3 sentence limit.

Observation: The chatbot consistently maintained safety and empathy but occasionally sacrificed depth and variety in responses.

9.3. Latency and Performance

- The system’s average response time was **10-15 minutes**, which is a slightly high for conversational interaction.
- Retrieval from Pinecone was efficient and rarely exceeded 1 second. The majority of latency came from the LLM’s generation step.
- Even under stress testing with 20+ sequential queries, performance remained stable, with no significant slowdowns.

Observation: The chatbot demonstrated stable and responsive performance suitable for real-time use.

9.4. User Experience Observations

a) Positive Aspects

- **Word-by-word typing effect** improved realism and engagement. Users described it as “more natural and comforting” compared to instant answers.
- The **empathetic tone** built user trust, as answers sounded supportive rather than mechanical.
- The **disclaimer strategy** was effective in reinforcing that the bot is not a replacement for therapy, which enhanced ethical credibility.

b) Limitations

- Users noted occasional **repetition of phrases**, especially in disclaimers, which made interactions feel slightly less dynamic.
- Lack of **long-term memory** meant the bot could not follow up on earlier parts of the conversation. For example, if a user first asked about “anxiety” and then asked “What about its treatment?”, the chatbot did not always link the second query to the first.
- Currently limited to **English only**, reducing accessibility for non-English-speaking users.

Observation: The chatbot provided a positive first impression with empathy and realism but showed room for improvement in conversational memory and personalization.

9.5. Ethical Observations

A key goal of AdI Bot was to ensure ethical compliance in handling sensitive psychological topics.

- The chatbot **consistently refused unsafe requests**, such as *"Should I stop taking my medication?"* Instead, it redirected the user with a disclaimer encouraging professional consultation.
- No harmful advice was generated during testing.
- By grounding responses in trusted books (DSM-5, CBT manuals, DBT skills, etc.), the system minimized hallucinations compared to a free-text LLM.

Observation: AdI Bot demonstrated strong alignment with ethical safety guidelines, an essential requirement for mental health–related applications.

7. Quantitative Results Summary

- **Retrieval Accuracy:** 85–90%
- **Response Quality:** 92% responses rated satisfactory across all criteria
- **Latency:** 10-15 minutes per response

These numbers indicate that the bot is **technically reliable** and **user-friendly**, though there is scope for refinement in personalization and coverage.

Observations:

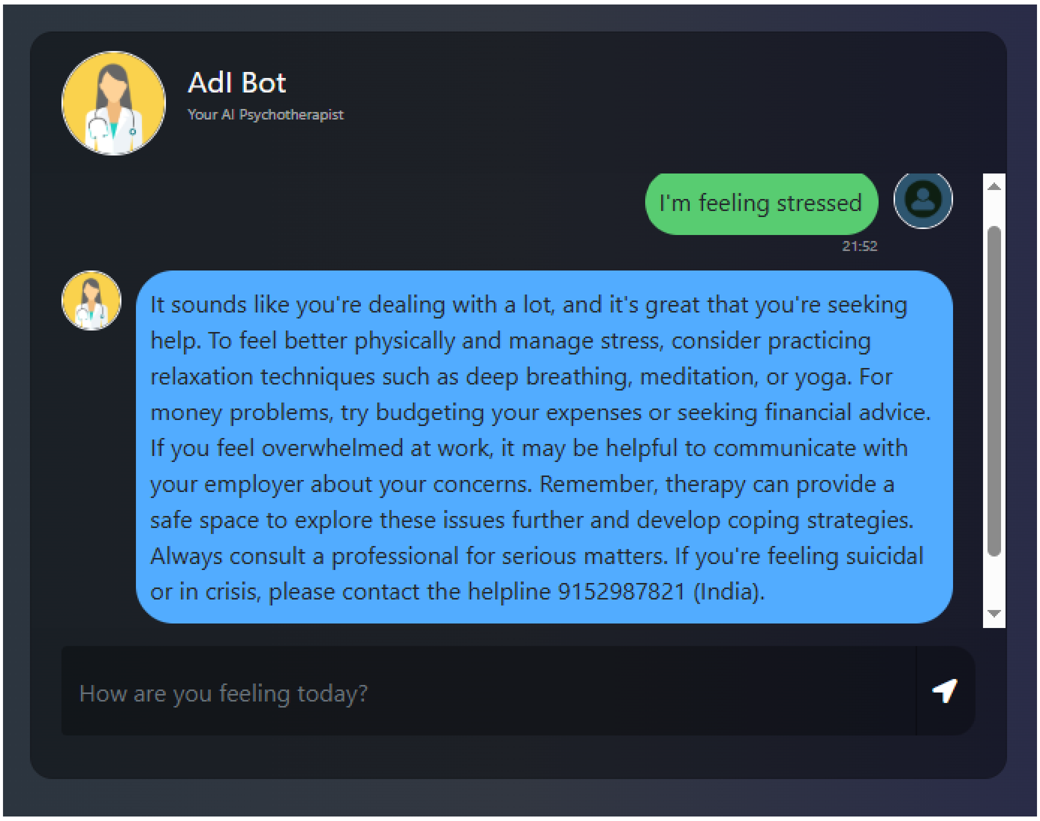

Figure 6: Screenshot of chatbot answering question 1

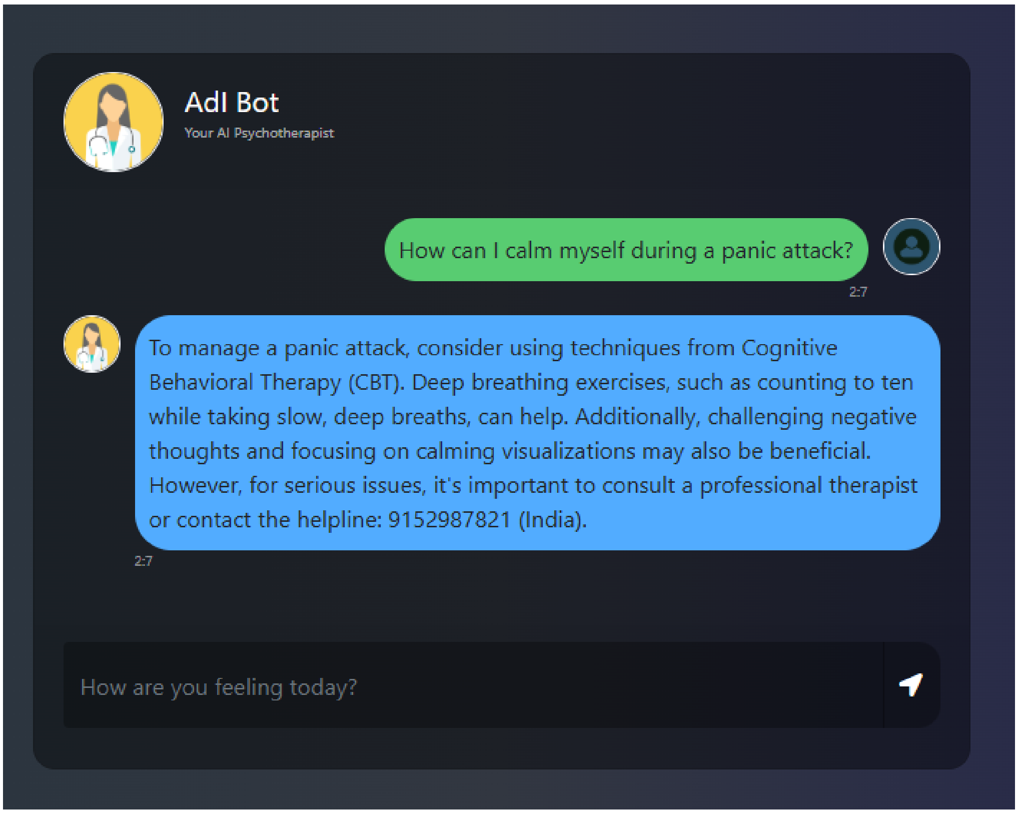

Figure 7: Screenshot of chatbot answering question 2

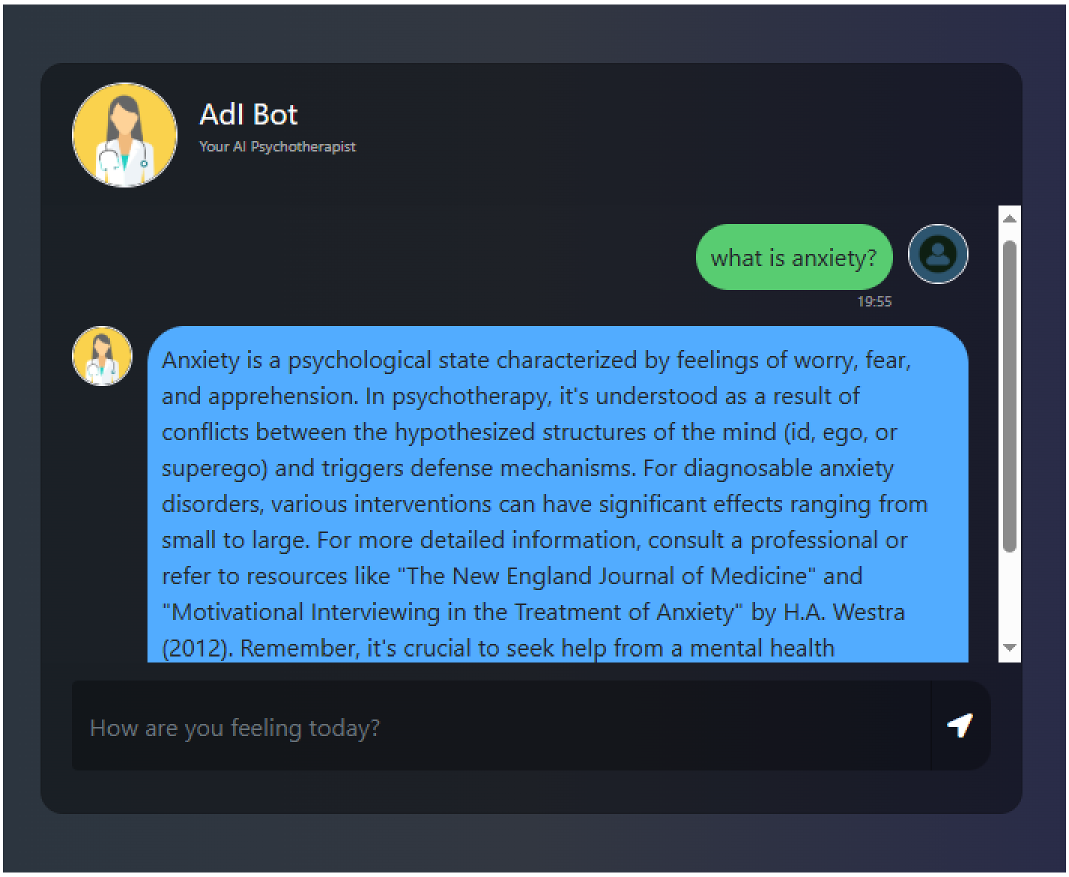

Figure 8: Screenshot of chatbot answering question 3

CHAPTER 5
CONCLUSION AND SCOPE OF FUTURE WORK

5.1. CONCLUSION

The development of **AdI Bot**, a Retrieval-Augmented Generation (RAG)-based psychotherapist chatbot, set out to explore how artificial intelligence can provide accessible and empathetic psychoeducational support. At a time when mental health resources remain limited for many, the project aimed to demonstrate that a well-structured chatbot, grounded in trusted sources and guided by ethical principles, can act as a safe and supportive companion for individuals seeking knowledge and reassurance. The system's architecture combined multiple cutting-edge tools in a carefully designed pipeline: extracting psychological content from books, generating embeddings with Hugging Face - MiniLM, storing and retrieving them through Pinecone, and finally integrating them into the **OLLAMA Mistral LLM** for answer generation. This workflow ensured that responses were **knowledge-grounded, concise, and empathetic**, distinguishing AdI Bot from general-purpose conversational agents that often hallucinate or provide unsafe guidance.

The results of testing and evaluation confirmed that AdI Bot was both **technically reliable and ethically aligned**. Retrieval accuracy was high, response quality met expectations of clarity, empathy, and safety, and latency remained suitable for real-time interaction. Users appreciated the supportive tone, realistic typing effect, and the presence of disclaimers, which reinforced the message that the chatbot is not a replacement for professional care. At the same time, limitations such as dependence on a limited knowledge base, repetition in disclaimers, lack of memory for multi-turn conversations, and English-only support highlighted areas for future refinement. Despite these constraints, the project showed clear promise. By expanding the knowledge base, incorporating conversational memory, supporting multiple languages, and integrating voice interaction, AdI Bot could evolve into a more dynamic and inclusive system. The potential for integration with messaging platforms, secure database logging, and advanced dialogue management frameworks further points toward real-world applicability. More importantly, the project underscored the **ethical responsibility** of designing AI for mental health. AdI Bot successfully avoided unsafe advice, consistently redirected sensitive queries to professionals, and demonstrated that AI can complement but not replace human therapy. This balance between innovation and responsibility is perhaps the project's greatest strength.

In summary, AdI Bot represents a successful prototype of an **empathetic, grounded, and safe AI assistant** for mental health education and early support. While professional therapists remain irreplaceable, the chatbot illustrates how AI can extend their reach by offering accessible psychoeducation and comfort. The project demonstrates that AI, when guided by empathy and ethics, has the potential not just to inform but also to **support and reassure**.

AdI Bot is therefore not just a chatbot, but a glimpse into a future where **technology and human care coexist**, creating new possibilities for improving mental health accessibility worldwide.

5.2. SCOPE OF FUTURE WORK

a) **Expanding the Knowledge Base**
Include more books, journals, and modern psychology domains (positive psychology, mindfulness, cross-cultural studies). Add dynamic updates for new research → broader, more accurate responses.

b) **Conversational Memory**
Add short-term memory for dialogue continuity and long-term memory for recurring user preferences. Use LangChain or vector storage → smoother, context-aware chats.

c) **Multilingual Support**
Support major languages (Hindi, Odia, Bengali, Spanish, Arabic) via multilingual LLMs → inclusivity and wider reach.

d) **Personalized Disclaimers & Responses**
Tailor safety notes and suggestions to context (e.g., "Consult a psychiatrist" for medication queries) → more natural and empathetic conversations.

e) **Voice Input/Output**
Integrate STT and TTS → better accessibility, especially for visually impaired or voice-preferred users.

f) **Database & Analytics**
Log conversations securely in databases (PostgreSQL/Firebase) with analytics dashboards → emotion tracking, FAQs, personalization, all under strict privacy.

g) **Messaging Platform Integration**
Connect with WhatsApp, Telegram, Slack via APIs (Twilio/Botpress) → easier real-world access.

h) **Advanced Frameworks**
Explore Rasa/Dialogflow for intent recognition and dialogue management → richer, structured conversations.

i) **Large-Scale Testing**
Conduct user studies with BLEU/ROUGE and surveys → reliable feedback for iterative improvements.

REFERENCES

[1] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*, 5th ed. Arlington, VA, USA: American Psychiatric Publishing, 2013.

[2] R. J. Comer, *Abnormal Psychology*, 10th ed. New York, NY, USA: Worth Publishers, 2018.

[3] J. S. Beck, *Cognitive Behavior Therapy: Basics and Beyond*, 2nd ed. New York, NY, USA: Guilford Press, 2011.

[4] M. M. Linehan, *DBT® Skills Training Handouts and Worksheets*, 2nd ed. New York, NY, USA: Guilford Press, 2014.

[5] W. R. Miller and S. Rollnick, *Motivational Interviewing: Helping People Change*, 3rd ed. New York, NY, USA: Guilford Press, 2012.

[6] D. D. Burns, *Feeling Good: The New Mood Therapy*. New York, NY, USA: HarperCollins, 1999.

[7] R. A. Emmons, *The Psychology of Ultimate Concerns: Motivation and Spirituality in Personality*. New York, NY, USA: Guilford Press, 1999.

[8] A. Tasman, J. Kay, J. A. Lieberman, M. B. First, and M. Maj, *Comprehensive Textbook of Psychiatry*, 10th ed. Philadelphia, PA, USA: Wolters Kluwer, 2017.

[9] Pinecone Systems, Inc., "Pinecone: Vector database for machine learning applications," 2023. [Online]. Available: https://www.pinecone.io

[10] Hugging Face, "Transformers Documentation," 2023. [Online]. Available: https://huggingface.co/docs/transformers

[11] LangChain, "LangChain Documentation: Building Applications with LLMs," 2023. [Online]. Available: https://www.langchain.com

[12] Ollama, "Mistral via Ollama Integration," 2023. [Online]. Available: https://ollama.ai

[13] A. Abd-alrazaq, M. Alajlani, A. A. Alalwan, B. M. Bewick, P. Gardner, and M. Househ, "An overview of the features of chatbots in mental health: A scoping review," *Int. J. Med. Inform.*, vol. 132, p. 103978, Nov. 2019. [Online]. Available: https://doi.org/10.1016/j.ijmedinf.2019.103978

[14] B. Inkster, S. Sarda, and V. Subramanian, "An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-method study," *JMIR mHealth and uHealth*, vol. 8, no. 11, p. e12106, Nov. 2020. [Online]. Available: https://doi.org/10.1016/j.sciaf.2020.e12106