MindSpark: AI Psychiatric Assistant

A Comparative Analysis of DeepSeek, Hybrid RAG, and Simple RAG Models

March 6, 2025

Team Members

2021-CS-63 Mutiaba Mohsin
2021-CS-67 Ammad Aslam
2021-CS-77 Uswa Arif
2021-CS-78 Fatima Muskan

Supervisor
Dr. Syed Khaldoon Khurshid

Introduction

The MindSpark project aims to develop an AI psychiatric assistant capable of diagnosing mental health disorders and providing treatment recommendations through an interactive chatbot. Given the constraints of limited training data, we opted for a **Hybrid RAG (Retrieval-Augmented Generation) Model**, which combines retrieval-based approaches with generative AI.

The Hybrid RAG model was selected because:

- **Data Limitations**: Due to the unavailability of large-scale structured psychiatric datasets, a purely generative model (like DeepSeek) could introduce hallucinated responses. Hybrid RAG ensures accuracy by retrieving existing medical knowledge before generating responses.
- **Better Control and Interpretability**: Unlike end-to-end generative models, our approach enables better control over the model's responses by restricting it to retrieved relevant content.
- **Empathetic and Context-Aware Conversations**: By leveraging LLaMA-3 API for response generation, Hybrid RAG ensures that retrieved questions are reformulated empathetically, enhancing user engagement.
- **Optimized Performance**: The model balances retrieval and generation, making it both computationally efficient and scalable for real-world deployment.

This document presents a comparative analysis of three AI models: **DeepSeek, Hybrid RAG, and Simple RAG**, evaluating their strengths, weaknesses, and suitability for our project.

Model Comparisons

DeepSeek Model

Overview: Uses OpenRouter's DeepSeek API for psychiatric consultations.

Strengths:

- Powerful LLM (DeepSeek-R1-Distill-8B).
- End-to-end automation for disorder detection and response generation.
- FAISS integration for efficient similarity search.

Weaknesses:

- Over-reliance on API calls.
- Limited RAG implementation.
- Less control over responses.

Output:

```
Detected Disorders: ['Generalized Anxiety Disorder', 'Panic Disorder']
Chat Summary: 'User is experiencing persistent anxiety and panic episodes.'
Questions Asked: ['Have you noticed any triggers for your anxiety?', 'How often do you
    experience panic attacks?']
```

Hybrid RAG Model

Overview: Combines retrieval and generation using Groq API (LLaMA-3) for empathetic responses.

Strengths:

- Better question generation with rephrasing.
- More control over AI responses.
- Faster and optimized API calls.

- Handles multi-turn conversations naturally.

Weaknesses:

- Requires an API key (Groq).
- Slightly more computationally intensive.

Output:

```
Detected Disorders: ['Generalized Anxiety Disorder', 'Panic Disorder']
Rephrased Question: "I understand that anxiety can be overwhelming. Can you tell me more
    about it?"
Questions Asked: ['When did you start experiencing these symptoms?', 'Do you experience
    panic attacks?']
```

Simple RAG Model

Overview: A lightweight retrieval-based model without LLM-generated responses.

Strengths:

- Fully local (no API dependency).
- Lightweight and efficient.
- Ensures dataset-restricted outputs.

Weaknesses:

- No AI-generated responses.
- Not as conversational.
- No rephrasing or contextualization.

Output:

```
Retrieved Questions: ['Do you frequently experience excessive worry?', 'Do you have
    difficulty falling asleep?']
Detected Disorders: ['Generalized Anxiety Disorder', 'Insomnia Disorder']
Questions Asked: ['Do you feel tired during the day?', 'Do you overthink situations
    frequently?']
```

Conclusion

- Use **Hybrid RAG Model** for the most accurate, empathetic, and efficient AI psychiatric assistant.
- Use **Simple RAG Model** for offline, resource-limited applications.
- Use **DeepSeek Model** only if full automation is required.