Pandora: AI Counselor Bot

Developed by:
Michelle De Melo (24P0620007)
Radhika Dabholkar (24P0620009**)**

1. Introduction

Mental health issues like stress, loneliness, anxiety, and depression are becoming increasingly common among students.
There is a strong need for solutions that can provide preliminary emotional support safely and empathetically.
Pandora is an AI-driven virtual counselor developed to engage in warm, supportive conversations by combining advanced retrieval techniques and model fine-tuning.

2. Problem Statement

Traditional chatbots often:

- Fail to handle emotional conversations carefully.
- Hallucinate and provide unsafe or inaccurate advice.
- Sound robotic and emotionally disconnected.

There is a clear requirement for an AI counselor that can:

- Respond with empathy.
- Ensure factual safety.
- Interact warmly with users facing emotional distress.

3. Approach

Pandora uses a hybrid approach:

- **Retrieval-Augmented Generation (RAG):**
  Instead of generating random replies, the system retrieves relevant, real-world counseling advice stored in a database.
- **Fine-Tuning of Transformer Models (Optional):**
  A flan-T5-small model was fine-tuned to rewrite retrieved advice in a softer, more empathetic tone.
- **Fallback Strategy:**
  If fine-tuning is not available, the system directly uses retrieved safe advice to ensure user safety.

4. System Design (Short Version)

Pandora uses a **layered architecture** to handle conversations:

- **Input Layer:**
  Takes the user's emotional or casual text input.
- **Embedding Layer:**
  Converts the input into a numerical format using the MiniLM-L6-v2 model.
- **RAG Layer:**
  Uses FAISS to search for the most relevant advice from a cleaned emotional support dataset.
- **(Optional) Fine-Tuned Generation Layer:**
  Improves the advice by making it more warm and personalized using a fine-tuned flan-t5-small model.
- **Response Layer:**
  Sends the final supportive message back to the user.

Key Components

- **Data Source:** intents_cleaned.json (emotional advice)

- **Embedding Model:** all-MiniLM-L6-v2
- **FAISS Index:** For fast advice retrieval
- **RAG Module:** Fetches the closest emotional advice
- **Fine-tuned Model (optional):** Warms up the response
- **Frontend:** Terminal interface (future: Streamlit app)

5. Technologies Used

- **Python 3.10 —** Programming language.
- **HuggingFace Transformers —** For loading and fine-tuning models.
- **SentenceTransformers (MiniLM-L6-v2) —** For sentence embeddings.
- **FAISS (Facebook AI Similarity Search) —** For fast retrieval of relevant advice.
- **Jupyter Notebook —** For coding and model training.
- **Streamlit (Optional) —** For building a simple user interface.

6. Results

Pandora successfully handled emotional queries by providing:

- **Warm and empathetic responses** for feelings like sadness, anxiety, and loneliness.
- **Factually safe advice**, especially for sensitive topics like depression.
- **Human-like conversations** without hallucinations.

Sample Results:

User Input	Pandora's Response
"I feel sad today."	"It's okay to feel sad sometimes. You're not alone. I'm here to listen."
"I'm anxious about exams."	"Anxiety is normal. Try breathing deeply and remember you're doing your best."
"I feel like ending it all."	"I'm sorry you're feeling this way. Please reach out to someone you trust or a mental health professional."

7. Challenges

- Ensuring **complete safety** of the chatbot's responses.
- Cleaning large datasets of emotional advice to remove unsafe content.
- **Fine-tuning models** using limited computational resources (small GPUs).
- Designing a **fallback system** to avoid unsafe generation even without fine-tuning.

8. Future Work

- Fine-tune **larger models** like Mistral-7B or Llama-2-7B-chat for better emotional intelligence.
- Build a full **web-based chat application** using Streamlit or ReactJS.
- Implement **emotion detection**, allowing responses to auto-adapt based on user emotions.
- Add **multilingual support** to allow chatting in Hindi, Spanish, and more languages.

9. Conclusion

Pandora shows that combining **retrieval techniques** with **fine-tuned AI models** creates a safe and supportive conversational agent.
The project successfully handled emotional conversations with warmth and factual accuracy.
With future improvements, Pandora can become an even more powerful emotional support tool for students and other users worldwide.