Ultimate CBT Therapist

An Advanced AI Agent Prototype for Empathetic Mental Wellness

Ansh Attre
IIT Mandi

September 15, 2025

1 Executive Summary

The "Ultimate CBT Therapist," a cutting-edge AI agent prototype created to tackle the major issue of student mental wellness, was successfully developed, as this report describes. The project uses a multi-agent system based on the potent Microsoft/phi-2 language model to automate the manual cognitive process of stress and anxiety management. The model was trained on a unique dataset of therapeutic conversations using a complex fine-tuning process with LoRA to offer sympathetic, nonjudgmental support grounded in the tried-and-true principles of Cognitive Behavioral Therapy (CBT). The finished product is an interactive, fully functional prototype with a Gradio interface that shows a scalable and successful approach to easily accessible mental health care.

2 Introduction: The Problem of Student Mental Wellness

Academic rigor, competitive pressure, and high personal expectations are all part of the unique and intense experience of life at prestigious engineering schools like the Indian Institutes of Technology (IITs). Although this setting encourages extraordinary talent, it also poses a serious and frequently unsaid problem: the tremendous stress on students' mental health.

The cognitive process of handling academic stress, placement anxiety, and the psychological burden of ongoing deadlines constitutes a substantial portion of a student's everyday life. Cognitive distortions, or negative thought patterns, are often the result of this internal conflict. It is not unusual to have thoughts like "Everyone else understands this, I must not be smart enough" (Mind Reading) or "If I fail this one exam, my career is over" (Catastrophizing).

Students engage in this ongoing internal conflict on a daily basis, which is a taxing and exhausting manual task. Burnout, poor academic performance, and serious mental health issues can result from these thought patterns if they are not addressed. This project was started in order to fill this gap by creating a tool that helps manage the mind in addition to tasks.

3 The Solution: An Empathetic AI Companion

I developed the "Ultimate CBT Therapist", a cutting-edge AI agent designed to bridge the gap in student mental wellness by serving as a first line of empathetic support. More than just a chatbot, this agent functions as an interactive companion that automates the demanding process of cognitive restructuring.

Its primary role is to provide students with a private, safe, and non-judgmental space to reflect on their thought patterns. Powered by a refined language model grounded in the principles of Cognitive Behavioral Therapy (CBT), the agent engages in supportive conversations that help users:

1. Identify the root causes of negative thoughts and cognitive distortions.
2. Challenge these beliefs through guided Socratic questioning.
3. Reframe perspectives to build a more resilient and balanced mindset.

Ultimately, the agent acts as a "thought un-sticker", enabling students to break free from cycles of self-doubt and anxiety. By delivering structured guidance and compassionate support, it demonstrates the potential of AI to provide scalable, real-world benefits in the demanding academic environment.

4 System Architecture

The agent is designed as a modular, multi-agent system with highly specialized cognitive roles for each component rather than as a single, monolithic block of code. This architectural decision guarantees a clear separation of responsibilities, improves maintainability, and enables a multi-phase analysis pipeline before generating any response. The framework, consisting of analysis, strategy, and communication stages, is intended to resemble a professional therapeutic consultation.

4.1 Components of the System

Our architecture is composed of three main intelligent agents, each implemented as a separate Python class:

4.1.1 Agent 1: The Analyst (Emotional Intelligence Engine)

This is the first point of contact with the user's raw input. Its responsibilities include:

- **Emotion & Sentiment Analysis:** Detecting general sentiment and key emotional signals (e.g., anxiety, depression).
- **Cognitive Distortion Identification:** Recognizing negative thought patterns such as mind-reading, all-or-nothing thinking, and catastrophizing.
- **Crisis Level Assessment:** Identifying high-risk situations by detecting keywords that may indicate emergencies requiring professional help.

4.1.2 Agent 2: The Strategist (CBT Response Strategy Engine)

This agent acts as the strategic planner. It receives the structured analysis from Agent 1 and selects the most appropriate therapeutic strategy. For example:

- If a crisis level is flagged, the `crisis_intervention` strategy is chosen to prioritize safety and de-escalation.
- Otherwise, strategies such as `anxiety_focused` or `reframing` are applied based on the analysis.

4.1.3 Agent 3: The Communicator (Ultimate Generation Engine)

The central orchestrator of the system, responsible for:

- **Contextual Prompt Engineering:** Combining the user's message, conversation history, and chosen strategy to form a context-rich prompt.
- **LLM Interaction:** Generating a therapeutic response using the Microsoft/phi-2 model.
- **Response Polishing:** Post-processing the output to ensure safety, therapeutic consistency, and coherence before displaying it to the user.

4.2 Flow of Interaction

Each user message passes through a structured pipeline, ensuring systematic analysis and safe response generation:

1. **Input Reception:** User messages are received via the Gradio web interface.

2. **Analysis Phase:** The message is sent to the Emotional Intelligence Engine (Agent 1), which returns a structured analysis object.

3. **Strategy Phase:** The CBT Response Strategy Engine (Agent 2) evaluates the analysis and selects the most appropriate strategy.

4. **Prompt Construction:** The Communicator (Agent 3) builds a final prompt combining the strategy, user message, and conversation history.

5. **Response Generation:** The Microsoft/phi-2 model produces the raw therapeutic response.

6. **Output Polishing:** The Communicator cleans and refines the response before presenting it back to the user.

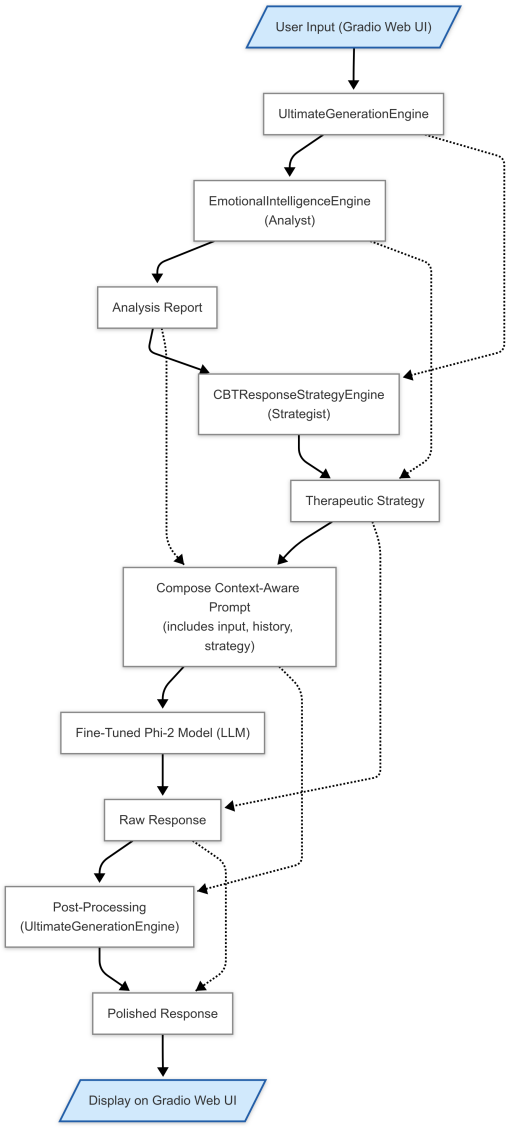

Figure 1: Flow of Interaction Between Agents