**Team 26: AI Enthusiasts**

-Priyadarshini Balasubramanian

-Sana Ambreen

-Sumanth Sudha

-Madhumitha Natesan

- Mohammad Mahmud

**Problem Statement**:

Create a mental health support chatbot using an open-source or hosted LLM that provides empathetic, ethically safe responses.

**Project Overview:**

This system provides empathetic, ethically safe, and context-aware conversational support for users experiencing emotional distress or seeking wellness guidance.  
It combines Dialogflow CX’s structured conversational design with Large Language Model (LLM) intelligence and speech-to-text / text-to-speech integration to ensure accessibility and natural interaction.

The chatbot’s design follows ethical mental-health AI principles — no diagnosis, no data collection, and immediate escalation for crisis cases.

**System Architecture:**

User (Voice/Text)  
↓  
Text/Speech-to-Text (Dialogflow Audio Input)  
↓  
Dialogflow CX Agent  
├── Playbooks (Welcome, Mood Assessment, Crisis, Supportive, Self-Help, Group Suggestion, Out-of-Scope, Closure)  
├── Tools (Empathetic Response Generator, Sentiment Analyzer, Hallucination Guard, Conversation Memory Manager)  
└── Random Therapy Event Generator (inline simulation)  
↓  
Text→ User

**Conversation design:**

**A screenshot of a diagram

AI-generated content may be incorrect.**

**Core Components:**

**Playbooks:**

| **Playbook** | **Function** |
| --- | --- |
| **Welcome & Disclaimer** | Greets user, clarifies that chatbot is not a therapist, and starts mood check. |
| **Mood Assessment** | Uses sentiment/emotion analyzer to classify emotional intensity (mild / moderate / high). |
| **Crisis Support** | Detects suicidal/self-harm cues and triggers an emergency resource message. |
| **Supportive Conversation** | Offers reflective, empathic, non-clinical dialogue. |
| **Self-Help & Education** | Provides tips, coping techniques, or random simulated wellness events. |
| **Support Group Suggestion** | Recommends community or therapy sessions (simulated data). |
| **Out-of-Scope** | Politely declines unsafe or medical requests. |
| **Closure** | Concludes with reassurance or next-step suggestion. |

**Integrated Tools:**

| **Tool** | **Description** |
| --- | --- |
| **Empathetic Response Generator** | Uses an LLM (Gemini-2.5-flash) to create emotionally intelligent messages. |
| **Sentiment / Emotion Analyzer** | Classifies emotion (happy, sad, anxious, angry) and intensity level for routing. |
| **Hallucination Guard** | Filters out unsafe, speculative, or false content using content-validation rules. |
| **Conversation Memory Manager** | Tracks dialogue context to avoid repetition and ensure continuity. |

**Speech Integration:**

Dialogflow utilizes Google Cloud’s Speech-to-Text (STT) API to convert voice input into text.

Dialogflow CX’s voice input pipeline uses:

Google Cloud Speech-to-Text relies on end-to-end neural models such as Conformer-based ASR (Automatic Speech Recognition) architectures.

**Ethical & Safety Design:**

| **Principle** | **Implementation** |
| --- | --- |
| **No diagnosis or medication suggestions** | Out-of-scope handler filters such prompts. |
| **Crisis escalation** | High-intensity emotion routes to Crisis Support Playbook. |
| **Fact validation** | Hallucination Guard ensures grounded responses. |
| **Repetition control** | Memory Manager blocks repeated empathy prompts. |
| **Privacy protection** | All logs anonymized; no user identifiers stored. |

**Example Flow Configuration (Dialogflow CX)**

**Flow: Mood Assessment**

* Intent: Detect user mood → route based on confidence/intensity.
* Condition Routes:
  + If sentiment\_score < -0.7 → Crisis Support
  + If -0.7 ≤ sentiment\_score < 0 → Supportive Conversation
  + Else → Self-Help Playbook

**Example Conversation:**

**Case 1 & 2:**

A screenshot of a chat

AI-generated content may be incorrect. A screenshot of a phone chat

AI-generated content may be incorrect.

(1) (2)

(1) The agent detects the user’s intention and produces relevant responses.

(2) The Agent also remembers previous prompts and continues the conversation accordingly.

**Case 3:** Extreme Suicidal Thoughts

A screenshot of a phone

AI-generated content may be incorrect.

In this case, the user has extreme suicidal thoughts. The agent responds with the helpline number.

**Case 4:** A normal conversation

A screenshot of a chat

AI-generated content may be incorrect.

The agent tries to build a conversation for any type of prompt given by the user, like “How many fins does a fan have.

**Case 5:** Negative Intent and Positive Emotion.

A screenshot of a chat

AI-generated content may be incorrect.

There is a negative intent, “I got diabetes,” and a positive emotion, “I am very happy”. In this case, the agent asks for more clarification.

**Case 6**: User shares happy news with the agent

A screenshot of a chat

AI-generated content may be incorrect.

The agent works as expected, following the path to analyze the mood of the user and then supporting the mood in a positive way, and also ending the flow with a closure response.

**Case 7**: User asks for medicines

**A screenshot of a chat

AI-generated content may be incorrect.**

In this case, the agent doesn’t provide any medications.

**Future Enhancements:**

* Expand multilingual and multimodal capabilities for inclusive global access
* Integrate clinician dashboard and wearable data for early intervention insights
* Connect verified therapy and wellness event APIs for real-time community support
* Deploy as a privacy-first Progressive Web App for continuous accessibility
* CHIRON redefines responsible emotional AI bridging empathy, ethics, and technology for safer mental health support