

## **Project Proposal:** FriendBot — An Emotionally Intelligent Conversational Agent for Relationship Advice

**Team:** Haoyuan Chen, Daniel Kim, Ferras Coulibaly, Taein Yi

### **1. Problem Being Addressed**

Navigating personal relationships — whether romantic, platonic, or familial — can be confusing, stressful, and emotionally taxing. Many individuals seek advice but may be reluctant to approach friends or professionals due to fear of judgment, privacy concerns, or lack of access. While search engines and social media offer tips, the responses are often generic, impersonal, or even harmful. There is a growing need for a private, emotionally supportive, and intelligent tool to help people reflect on and navigate their relationship challenges in a human-centered way. This motivates the need for a personal “Friend” capable of emotionally intelligent advice, or just a kind, empathetic ear. We seek to develop one such “Friend” specialized in **Romantic Relationships**.

### **2. Motivation**

- Privacy & Accessibility: Not everyone feels comfortable talking to friends, therapists, or strangers on forums about sensitive relationship issues.
- Emotional Intelligence: Users want to feel heard and understood, not just receive facts or automated replies.
- Conversational Support: A thoughtful, nonjudgmental companion can help users process emotions, gain clarity, and feel less alone.
- Gap in Current Tools: Most chatbots or advice engines are scripted, overly clinical, or lack warmth and nuance.

### **3. Desired Effect**

#### **FriendBot should:**

- Engage users in natural, emotionally supportive conversations about relationship issues.
- Detect emotional tone and conversational intent (e.g., venting, seeking advice, looking for affirmation).

- Offer non-directive guidance, helpful reframes, and thoughtful questions.
- Maintain a consistent and empathetic personality that builds user trust over time.
- Avoid judgment, diagnosis, or scripted “fix-it” responses — focusing instead on companionship and perspective.

## 4. State of the Art

- Therapeutic Chatbots: Woebot, Replika, and Wysa incorporate CBT techniques but are limited in open-ended relationship dialogue.
- LLMs like GPT-4: Capable of generating nuanced, emotionally aware responses but may hallucinate or respond inappropriately without constraints.
- Prompt Engineering: Persona shaping and style conditioning enable more controlled, emotionally aligned output.
- Sentiment & Emotion Detection: NLP models like RoBERTa or BERT fine-tuned on emotional datasets (**GoEmotions**, EmpatheticDialogues) help interpret user affect.
- Memory & Contextual Coherence: Tools like LangChain enable coherent multi-turn dialogue and personalized context retention.

## 5. Proposed Method / Architecture

We propose a **Three layer Architecture** for processing User inputs instead of a traditional one turn pass through a Language Model like ChatGPT. This process is described below:

User input gets passed through the following layers:

1. Fine-tuned BERT Model (GoEmotions Dataset) to extract emotion
2. Openai GPT o3-mini to extract intent
3. Openai GPT 4o for final response

Details on the procedure is described in the following sections:

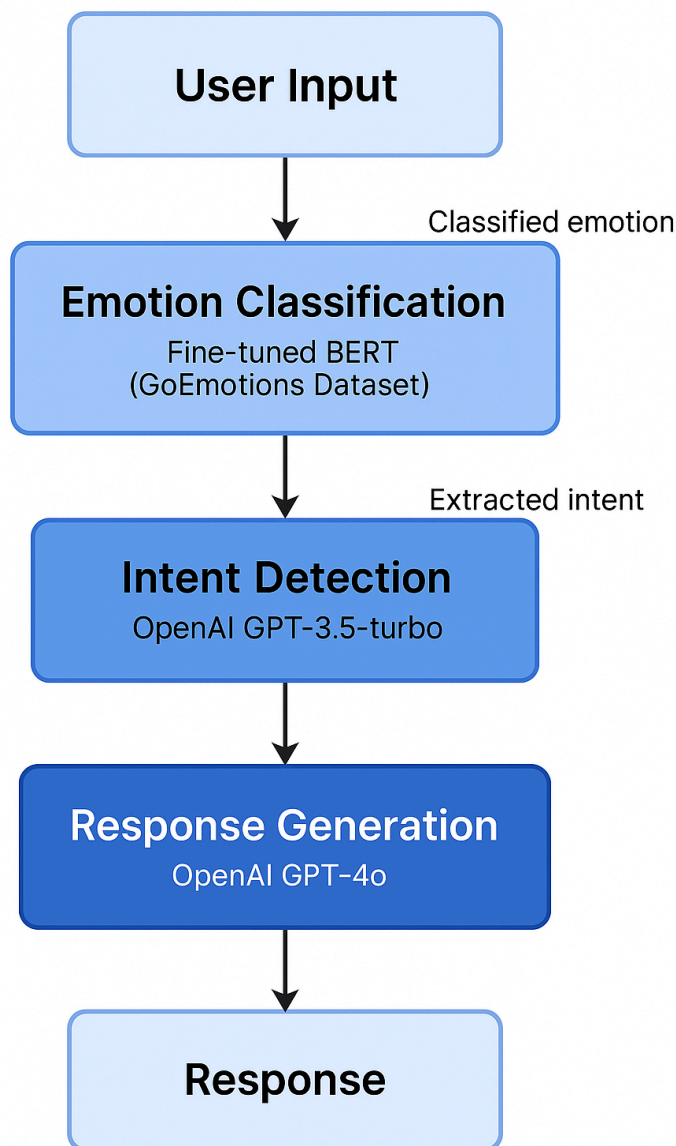
### A. User Input Analysis

- Classify message emotion (e.g., sadness, frustration, confusion, hope).

We achieve this by classifying using a **fine-tuned BERT Model (GoEmotions dataset)**. Based on the GoEmotions dataset, we discover a fine-tuned BERT Model or train one ourselves to accurately classify inputs into various emotions. This emotion factor is then used later on for improved FriendBot response generation.

- Detect intent: venting, seeking advice, reflection, etc.

This is done through a prompt-engineered Openai GPT o3-mini model. This will extract the intent behind the user input.



## **B. Empathetic Prompt Engineering**

- Incorporate detected emotion and intent to condition the LLM prompt:

We will develop a strong GPT prompt engineered towards a consistent, emotionally intelligent persona, which through interpreting the user's input, potential context, and detected factors (emotions and intent), will produce a suitable response as a "Friend" giving relationship advice.

A foundational / basic version can be as follows:

"You're FriendBot, a kind, emotionally intelligent companion. Your friend is feeling {emotion} and wants to talk about {topic}. Listen closely and respond with warmth, reflection, and thoughtful questions."

## **C. Response Generation with LLM**

- Use GPT-3.5 / GPT-4o or LLaMA-based models to generate:
- Reassurance ("That sounds really difficult.")
- Reframing ("Have you thought about it this way?")
- Insightful questions ("What do you want from this relationship?")
- Cautious suggestions ("If you're feeling unsafe or disrespected, it might be worth stepping back.")

## **D. Conversational Memory**

- Store key user context (e.g., past messages, names, emotional patterns).
- Refer back naturally: "Last time you mentioned you weren't sure if he was listening—how have things been since then?"

## **6. Feasibility**

- Open-source tools (Transformers, LangChain, Streamlit) make rapid prototyping viable.
- LLMs available via OpenAI API, Hugging Face, or local deployment.
- prompt tuning, retrieval pipelines, and chat UIs (from FinGPT Search Agent).

- MVP can focus on 3–4 core relationship themes: trust, communication, boundaries, and breakups.

- Dataset for emotional tone classification: GoEmotions, EmpatheticDialogues

## **8. Benchmark / Evaluation**

The benchmark will be split into three sections:

### **Section 1 (GoEmotions BERT Model):**

Using a set of test data (size TBD) with a similar format of the one below, we will run the inputs through the BERT model to extract emotions. We will then compare the detected emotions with the expected emotions and fine-tune to achieve 90% accuracy.

### **Section 2 (Openai GPT 03-mini Model):**

Using a set of test data (size TBD) with a similar format of the one below, we will run the inputs through the o3-mini model to extract intent. We will then compare the detected intent with the expected intent and prompt-engineer to achieve 90% accuracy.

### **Section 3 (Openai GPT 4o response with detected factors):**

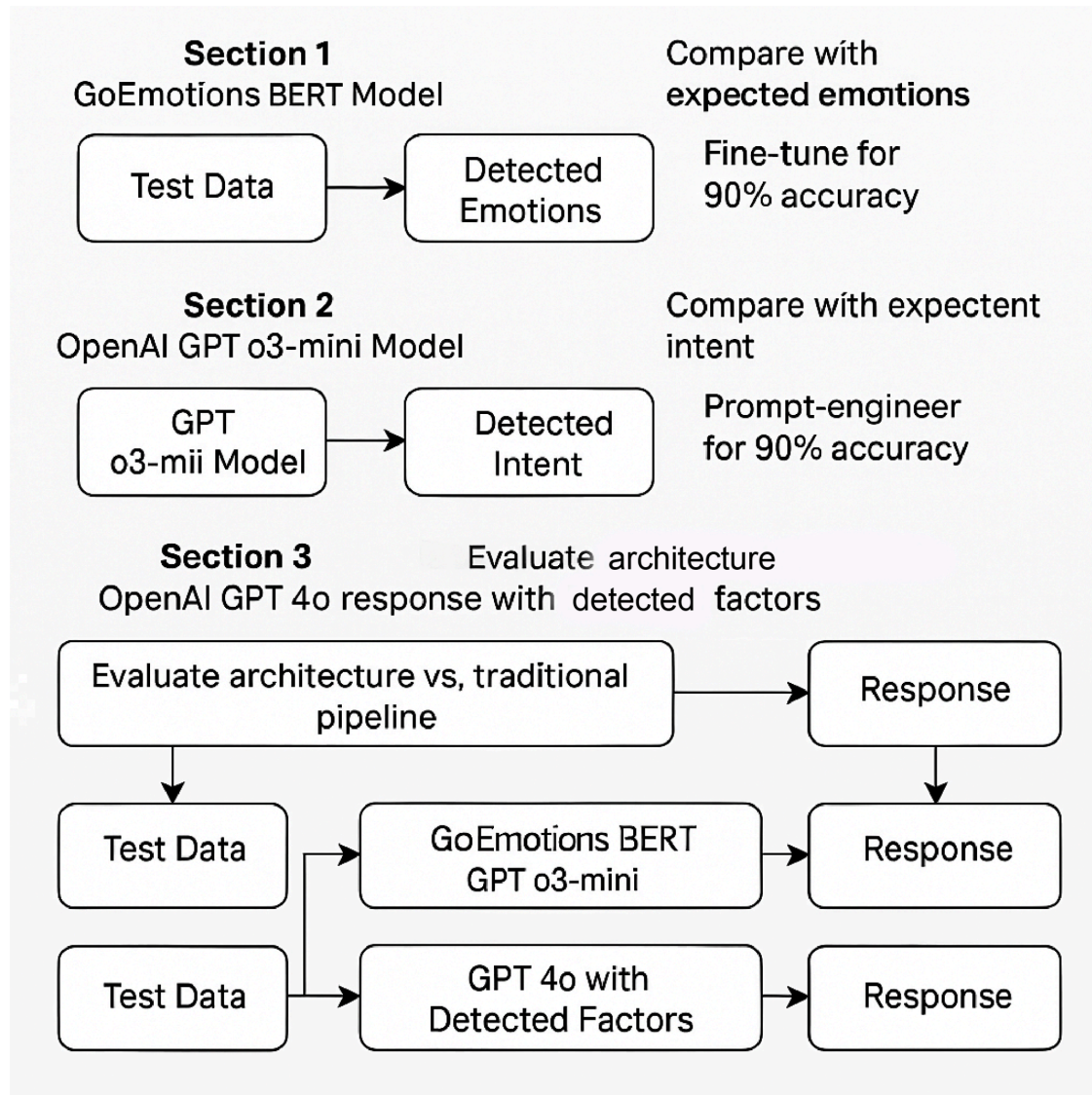
To evaluate our architecture against a traditional one-layer GPT input to output pipeline, we will conduct the following analysis.

First we run the inputs from another set of test data through GPT 4o only, so:

Input  $\Rightarrow$  GPT 4o  $\Rightarrow$  Response (Output)

Then we run the same inputs through our 3 layer architecture:

Input  $\Rightarrow$  GoEmotions BERT  $\Rightarrow$  GPT o3-mini  $\Rightarrow$  GPT 4o with the input and detected factors  $\Rightarrow$  Response (Output)



benchmark/evaluation Diagram

And we compare the results by the following potential steps:

1. Human Survey: After using the chatbots real users will complete a survey to assess each bot's ability to show empathy, reassurance, and maintain contextual awareness.
2. Automated Evaluation: Using an NLP pipeline, we can evaluate each chatbot's responses for key phrases that signal empathy, reassurance, and contextual relevance. From these evaluations, we will generate scores for these metrics that will go up and down based on how well each response follows these metrics.
3. Calibration: From the human surveys we can have an estimated score of what each chatbot should get and use that to calibrate metrics so that the automated score will be more accurate over time.

### Sample Test Dataset Section 1:

```
{  
  
  "user_input": "I feel like my partner doesn't really listen to me anymore.",  
  
  "expected_emotions": ["sad", "frustrated"]  
  
}
```

### Sample Test Dataset Section 2:

```
{  
  
  "user_input": "I feel like my partner doesn't really listen to me anymore.",  
  
  "expected_intent": ["seeking_advice"]  
  
}
```

Potential Datasets:

EmpatheticDialogues , DailyDialog, Persona-Chat, IEMOCAP

## 8. References

- Rashkin et al. (2019). Towards Empathetic Open-domain Conversation Models. ACL (<https://arxiv.org/pdf/1811.00207>)
- Demszky et al. (2020). GoEmotions: A Dataset of Fine-Grained Emotions. Google Research (<https://arxiv.org/pdf/2005.00547>)
- Wolf et al. (2020). Transformers: State-of-the-Art Natural Language Processing. Hugging Face (<https://aclanthology.org/2020.emnlp-demos.6.pdf>)
- OpenAI. (2023). ChatGPT: Optimizing Language Models for Dialogue
- Bickmore & Cassell. Relational Agents: A Model and Implementation of Building User TrustRelationshipBOT (<https://www.ccs.neu.edu/home/bickmore/publications/CHI2001.pdf>)