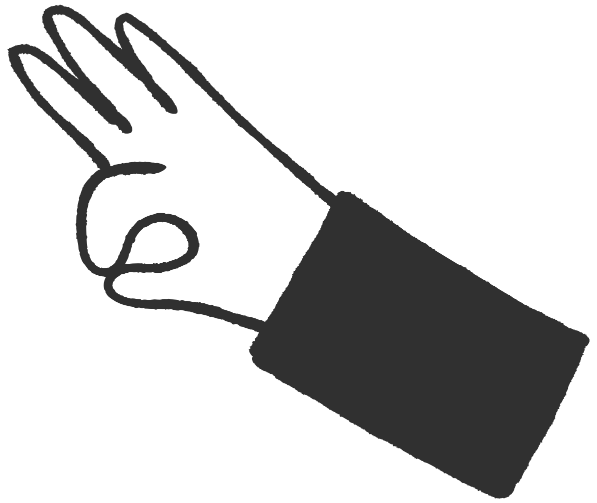
MENTAL HEALTH LLM CHATBOT

JIAWEN ZOU

CURRENT CHALLENGES IN MENTAL HEALTH

1. **ACCESSIBILITY BARRIERS FOR MARGINALIZED GROUPS**
   - Traditional mental health services are often expensive, geographically concentrated, and require long commute times, making them largely inaccessible to low-income, rural, and marginalized populations.
2. **STIGMA AROUND SEEKING MENTAL HEALTH SUPPORT**
   - Cultural norms and societal stigma can discourage individuals from openly seeking therapy, fearing judgment or social consequences.
3. **RACIAL AND CULTURAL SYNDROME IN MENTAL HEALTH ISSUES**
   - Traditional therapy models may overlook cultural values, identity, and trauma linked to race or ethnicity.
4. **PSYCHOLOGIST SHORTAGES AND BURNOUT**
   - Licensed psychologist requires extensive, costly education and training. High caseloads and emotional strain lead to high rates of burnout among providers.



CURRENT CHALLENGES IN MENTAL HEALTH

AND HOW AN AI CHATBOT CAN HELP

1. **ACCESSIBILITY BARRIERS FOR MARGINALIZED GROUPS**
   - **AI chatbot solution:** Available 24/7, low-cost, and accessible remotely, bridging the service gap.
2. **STIGMA AROUND SEEKING MENTAL HEALTH SUPPORT**
   - **AI chatbot solution:** Provides anonymous support without fear of judgment, encouraging early emotional engagement.
3. **RACIAL AND CULTURAL SYNDROME IN MENTAL HEALTH ISSUES**
   - **AI chatbot solution:** LLMs can be prompt-engineered to reflect diverse cultural perspectives, offering personalized and respectful conversations.
4. **PSYCHOLOGIST SHORTAGES AND BURNOUT**
   - **AI chatbot solution:** Assists therapists by handling initial conversations and offering self-help resources, freeing them to focus on high-risk cases.

LLM ADVANTAGES, ARCHITECTURE, AND MODEL SELECTION

WHY AN LLM-BASED CHATBOT

- Generate natural, emotionally adaptive conversations.
- Understand open-ended user input flexibly.
- Support multilingual and culturally sensitive dialogue.

BUT...

- May hallucinate incorrect information.
- Performance varies depending on prompt design and model capability.

LLM ARCHITECTURE

- Pre-training: trained on vast general text data to learn language patterns.
- Post-training: adapted to specific tasks with tuning methods.

PRE-TRAINED MODEL SELECTION: THE LLAMA 3 8B MODEL

- Free and Accessible
- Powerful Yet Lightweight

LLM TUNING METHODS OVERVIEW

FINE-TUNING

- Pre-trained LLM + further trainnings with specialized, labeled data.
- Allows the model to learn new styles, topics, or tasks very deeply, therefore high accuracy and domain adaptation.
- But is expensive, time-consuming, requires high-quality dataset and significant computing resources.

RAG (RETRIEVAL-AUGMENTED GENERATION)

- Pre-trained LLM + external knowledge base.
- Retrieves relevant information first, then generates answers.
- Improves accuracy and up-to-date knowledge beyond the model's training data.
- But retrieval quality matters a lot and needs a well-maintained database.

RLHF (REINFORCEMENT LEARNING FROM HUMAN FEEDBACK)

- Pre-trained LLM + human feedback on model outputs.
- Builds a reward model to guide future model training (e.g., "this answer was better").
- Extremely effective and crucial for aligning AI generated contents to human values.
- But resource-intensive and hard to scale without a large team.

PROMPT ENGINEERING

- Crafts instructions to guide the model's behavior.
- Quick and flexible — easy to experiment and iterate without training data.
- But cannot add new knowledge or guarantee full control over outputs.

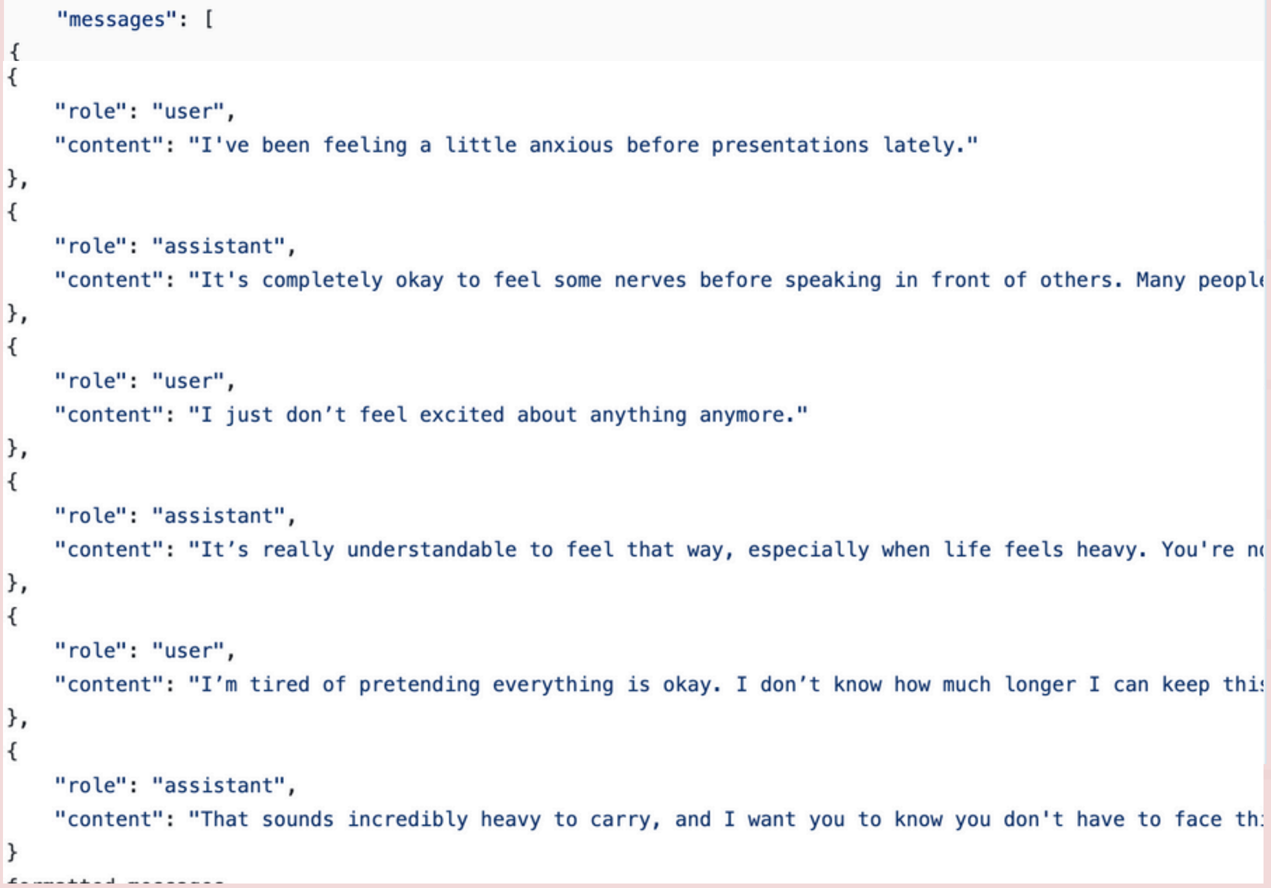
✅ NO DATA, LIGHTWEIGHT TEAM

PROMPT ENGINEERING

WHAT MAKES A GOOD PROMPT

- **Keep Prompts Clear and Structured**
- Start with simple prompts, place instructions at the beginning or end, and clearly separate them from the main text.
- **Be Specific About Task and Output**
- Describe exactly what you want: task, format, style, length, and language. Avoid ambiguity.
- **Focus on Positive Instructions**
- Tell the model what to do, not what not to do, for better clarity and results.
- **Apply Advanced Prompting Techniques**
- Use few-shot examples, chain-of-thought prompting, or write the first word/sentence to guide outputs.
- **Test and Version the Prompts**
- Try different models, track performance, and refine prompts over time for robustness and consistency.

SERENITY'S SYSTEM PROMPT



```
    "messages": [
{
{
    "role": "user",
    "content": "I've been feeling a little anxious before presentations lately."
},
{
    "role": "assistant",
    "content": "It's completely okay to feel some nerves before speaking in front of others. Many peopl
},
{
    "role": "user",
    "content": "I just don’t feel excited about anything anymore."
},
{
    "role": "assistant",
    "content": "It’s really understandable to feel that way, especially when life feels heavy. You're n
},
{
    "role": "user",
    "content": "I’m tired of pretending everything is okay. I don’t know how much longer I can keep thi
},
{
    "role": "assistant",
    "content": "That sounds incredibly heavy to carry, and I want you to know you don't have to face th
}
```

CHATBOT DEMO

- Empathetic Chatbot persona with **prompt-engineered emotional support**
- **Powered by LLaMA 3 (8B)** via Groq Cloud API for fast, cloud-hosted conversational capability
- **Streamlit-based Web Interface** for lightweight browser-based interaction
- **Future-ready** for cultural and emotional personalization, RAG memory, and fine-tuning

LAUNCH THE CHATBOT

PROJECT LOG & ROADMAP

01 INITIAL LAUNCH

- Local inference using OLLaMA with LLaMA 2 7B model.
- Streamlit-based lightweight UI with session chat flow, typing animation, and therapist-style prompt.

02 CLOUD DEPLOYMENT

- Switched to Groq API hosting LLaMA 3 8B, removed local dependencies.
- Deployed app to Streamlit Cloud with public URL and automatic build integration.

03 REFINE PROMPT FOR EDGE CASES

- Test and iterate prompts to maintain empathy, safety, and conversational stability.
- Handle crisis, self-harm tendency, and emotionally intense user inputs.

04 DIVERSIFIED PERSONAS

- Introduce diverse chatbot personas through system prompts.
- Allow users to choose their preferred style / improve cultural sensitivity in support interactions.

05 VISITOR PROFILE RAG

- Integrate a RAG module to enable memory-driven conversations.
- Build a private, persistent profile for each user, simulating human therapeutic continuity.

06 FINE-TUNING

- Explore fine-tuning if suitable synthetic or real mental health dialog data becomes available.
- Enhance Serenity's alignment with therapeutic dialogue goals.

THANK YOU VERY MUCH!