# A Fine-Tuned Large Language Model for Empathetic Mental Health Chatbot

# Overview

1. Problem
2. Objective
3. Overview
4. Dataset
5. Architecture
6. Evaluation
7. Results
8. Future Work

# Problem Statement

**Over 970 million people suffer from mental health disorders**

**Global shortage of over 4 million mental health professionals**

**Barriers:**

- **High cost of therapy**
- **Stigma surrounding mental health**
- **Limited geographic access**

**Need for scalable, empathetic, and accessible AI support tools**

# Project Objective

- **Build a web-based therapy chatbot using fine-tuned Large Language Models (LLMs)**
- **Deliver empathetic, CBT-aligned responses**
- **Detect user emotions with high accuracy**
- **Compare performance of 3 LLMs**
- **Ensure ethical, privacy-preserving, and user-centered design**

# System Overview

- **Fine-tuned LLMs:**
- **Llama-3.1-8B-Instruct**
- **Qwen/Qwen2-7B-Instruct**
- **Mistral-7B-Instruct-v0.3**

**Hybrid architecture:**

- **CBT-based rule prompts + generative replies**

# Dataset Preparation

- 20,000-row subset from 800,000-row mental health dataset
- Sources: public forums + synthetic dialogue generation
- Emotions covered: anxiety, grief, stress, loneliness, anger

Preprocessing steps:

- Cleaning, normalization, spelling correction
- Tokenization with Hugging Face tools

- 80/20 train-validation split with stratified sampling

# Fine-Tuning Process

- **Training environment: Google Colab T4 GPU**
- **Technique: Low-Rank Adaptation (LoRA) for memory efficiency**
- **Framework: Hugging Face Transformers**
- **Objective: generate CBT-aligned, emotionally appropriate replies**
- **Batch size: 8 (2 per device × 4 accumulation)**
- **Mixed-precision training with FP16**

# Hyperparameter Tuning

**Grid search over:**

- **Learning rates: 1e-5, 2e-4, 5e-4**
- **LoRA ranks: 8, 16, 32**
- **Batch sizes and accumulation steps**

**Best configuration:**

- **Learning rate = 2e-4**
- **LoRA rank = 16**
- **Dropout = 0.1**
- **Early stopping after 500 steps of no improvement**

# System Architecture

1. **Emotion Detection:**
   - Softmax classifier (5 classes: happy, sad, anxious, angry, neutral)
   - 90% accuracy using fine-tuned Qwen embeddings

2. **Dialogue Manager:**
   - Rule-based CBT prompts
   - LLM-generated empathetic responses
   - State tracking via Finite State Machine

# Evaluation Protocol

- **Evaluated using Gemini API on 100 test prompts**

**Criteria:**

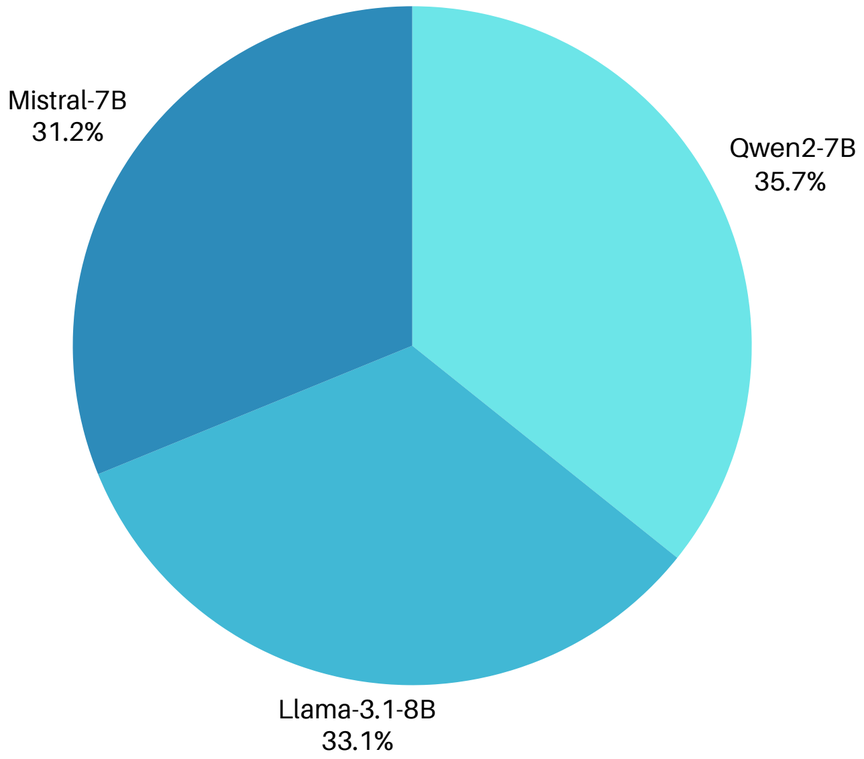
- **Empathy**
- **Understanding**
- **Advice Quality**
- **Conversation Flow**

**Judging function compared 3 model responses per prompt**

**Selected winner based on aggregated scores**

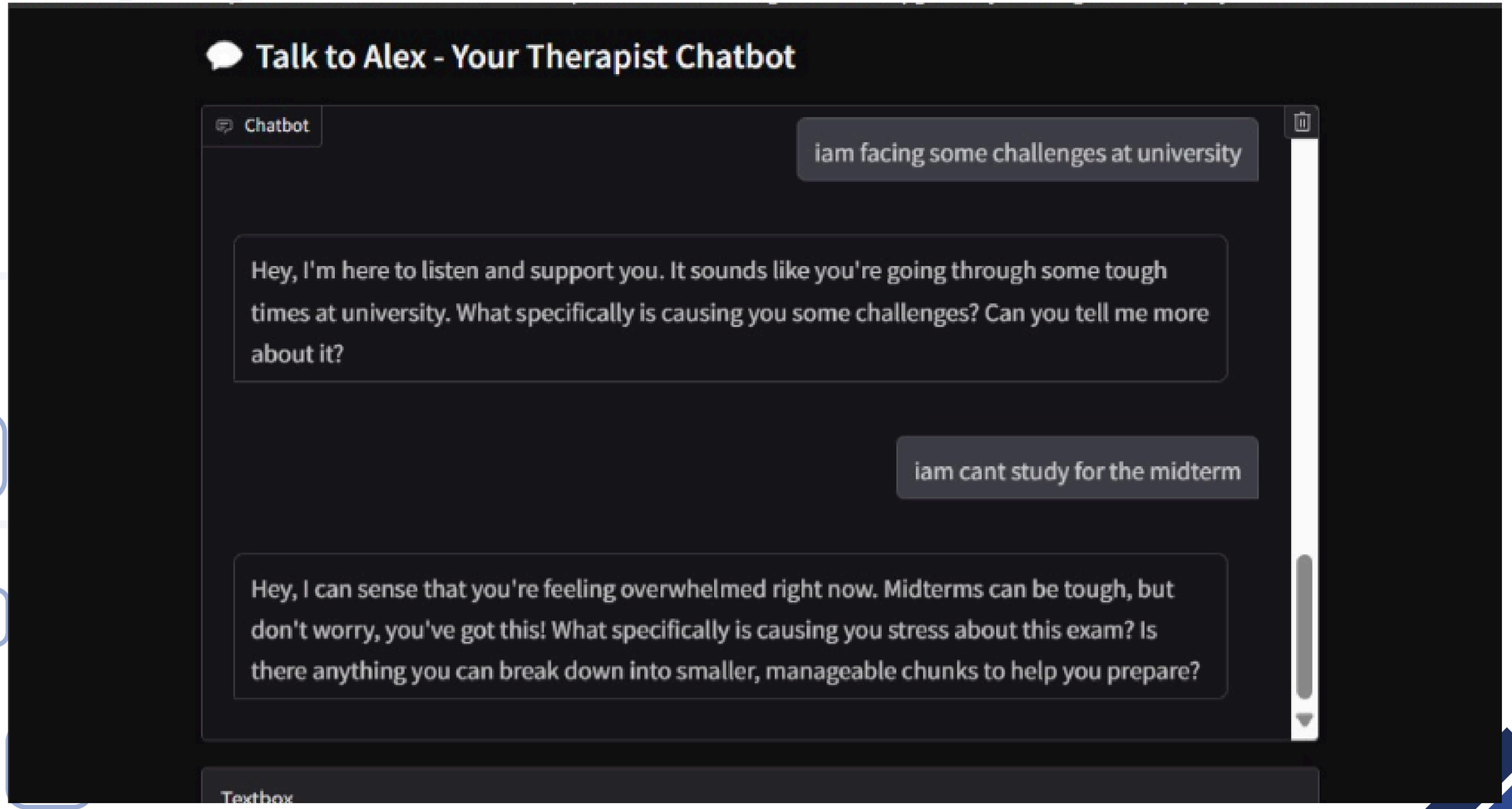
# Results - Fine-Tuning & Evaluation

- **Validation Loss (Best Checkpoint):**
- **Qwen2-7B: 0.8510**
- **Mistral-7B: 0.7352**
- **Llama-3.1-8B: 0.7241**




| Model | Empathy | Understanding | Advice | Flow | Overall |
|---|---|---|---|---|---|
| **Qwen2-7B** | 94% | 93% | 90% | 91% | 92% |
| Llama-3.1-8B | 87% | 86% | 84% | 83% | 85% |
| Mistral-7B | 82% | 80% | 79% | 78% | 80% |

# Demo


Talk to Alex - Your Therapist Chatbot
Chatbot
iam facing some challenges at university
Hey, I'm here to listen and support you. It sounds like you're going through some tough times at university. What specifically is causing you some challenges? Can you tell me more about it?
iam cant study for the midterm
Hey, I can sense that you're feeling overwhelmed right now. Midterms can be tough, but don't worry, you've got this! What specifically is causing you stress about this exam? Is there anything you can break down into smaller, manageable chunks to help you prepare?
Textbox

# Limitations

- **Dataset: only 20k rows used due to GPU constraints**
- **Struggles with complex emotional input (e.g., sarcasm)**
- **No real-time crisis detection system**
- **Internet dependency limits reach**
- **Gemini API sometimes misjudges nuance in emotional tone**

# Future Work

- **Train on full 800k dataset**
- **Add multilingual and offline support**
- **Integrate crisis detection and escalation protocols**
- **Combine generative responses with structured CBT modules**
- **Conduct clinical validation via longitudinal studies**

# Thank You