**DevelopersHub AI/ML Engineering Interns**

**DevHub Task # 05**

**Name: Muhammad Oun Haider**

**ID: DHC-1652**

**Task 5: Mental Health Support Chatbot (Fine-Tuned)**

**Abstract:**

This project developed an AI-powered mental health support chatbot using fine-tuned language models. The system provides empathetic responses to users experiencing stress, anxiety, or emotional distress. Leveraging the DistilGPT2 model fine-tuned on Facebook's Empathetic Dialogues dataset, the chatbot incorporates safety protocols and response enhancement mechanisms. The hybrid implementation combines neural network-generated responses with rule-based empathetic patterns, achieving functional emotional support capabilities while maintaining ethical safeguards against harmful advice.

**Introduction:**

Mental health challenges like stress and anxiety affect millions globally, yet access to support remains limited. This project addresses this gap by creating an accessible AI assistant that offers immediate emotional support. Unlike general chatbots, this system specializes in recognizing and responding to mental health cues using natural language processing. Built on transformer architecture, the chatbot demonstrates how AI can provide preliminary emotional support while maintaining ethical boundaries and directing users to professional resources when needed.

**Objectives:**

* Develop a fine-tuned language model capable of generating emotionally supportive responses
* Implement safety mechanisms for crisis situations
* Create a hybrid system combining neural network outputs with rule-based empathy patterns
* Design an intuitive interface for user interaction
* Ensure responses avoid harmful advice while validating user emotions
* Maintain computational efficiency for accessibility on consumer hardware

**Procedure:**

The development followed a structured pipeline: First, Facebook's Empathetic Dialogues dataset was preprocessed into "USER-BOT" conversation pairs with emotional context labels. The DistilGPT2 model was fine-tuned using PyTorch with a causal language modeling objective over 5 epochs. A post-processing enhancement layer was added to refine responses through:

* Safety filters detecting crisis keywords
* Empathy boosters adding validating phrases
* Diversity injectors preventing repetitive outputs.

The system was evaluated through iterative testing with mental health scenarios, measuring response relevance through keyword matching and qualitative assessment. The final implementation features a command-line interface with real-time response generation.

**Discussion:**

The chatbot achieved moderate success in generating contextually relevant responses, with particular strength in handling stress and anxiety scenarios. Key observations include:

* The fine-tuned model learned emotional language patterns but occasionally produced generic response
* Response enhancement significantly improved empathy and relevance (e.g., adding "That sounds difficult" to neutral outputs)
* Safety protocols effectively intercepted high-risk phrases like "self-harm" with crisis resources
* Computational efficiency allowed operation on consumer hardware (average response time <2s)
* Main challenges included dataset noise, over-repetition in responses, and balancing empathy with advice limitations. Future iterations would benefit from larger emotion-labeled datasets and clinician-reviewed response templates.

**Conclusion:**

This project demonstrates the feasibility of AI-assisted mental health support through empathetic chatbots. While not a replacement for professional care, the system provides accessible preliminary support that validates user emotions and offers constructive coping suggestions. The hybrid approach of combining neural networks with rule-based enhancements proved effective for maintaining response quality and safety. With further refinement in emotional intelligence training and crisis resource integration, such systems could meaningfully complement mental health ecosystems, particularly in resource-limited settings. The implementation successfully balances technological innovation with ethical responsibility in digital mental health support.