CAPSTONE PROJECT BRIEF

Title: Proof of Concept for Offline AI-Assisted Health Diagnosis

Background

Your startup aims to provide **offline health diagnostics** to rural communities, where rapid, accurate diagnosis can be lifesaving. As the newly hired AI Engineer, you are tasked with evaluating the feasibility of fine-tuning a **Mistral 7B** model for radiology-related question answering.

The Lead Engineer has suggested running a **Proof of Concept (PoC)** to determine whether fine-tuning significantly improves performance before allocating resources for full-scale development.

Objectives

1. Assess Mistral 7B's baseline performance on radiology-specific queries.
2. Fine-tune Mistral 7B using a curated radiology dataset.
3. Compare **pre- and post-fine-tuning** model performance.
4. Determine whether the fine-tuned model is a justifiable improvement for production use.

Key Requirements

- Open Source Tools Only – Use only freely available, open-source tools for model handling, fine-tuning, and evaluation.
- Fine-Tuning Dataset is **"belgiumhorse/share_gpt_style_patient_radiologist_data"** dataset hosted on Hugging Face.

3**. Model Deployment**

- Push the fine-tuned model to **Hugging Face Hub**.
- Deploy the same model locally using **Ollama** for inference.

4. **Web App Integration**

- Write a function leveraging Ollama APIs for local inference.
- **Include a code snippet and screenshot** of the function working as part of the PoC deliverable.

5. **Evaluation Procedure**

- Before fine-tuning, ask Mistral 7B four radiology-related questions and record its answers.
- After fine-tuning, repeat the same four questions with the new model.
- Compare the responses qualitatively (and quantitatively if feasible).

6. **Reporting**

Draft a management report detailing:

1. Approach & methodology
2. Performance comparison (before vs. after fine-tuning)
3. Cost/benefit considerations (compute, maintenance, accuracy)
4. Recommendation: Proceed with fine-tuning or use the base model

Suggested Deliverables

Baseline Evaluation Results (screenshots or text of answers)
Fine-Tuning Logs & Metrics
Link to Fine-Tuned Model on Hugging Face
Ollama Integration Function (code + evidence of successful inference)
Comparison Table (Before vs After Fine-Tuning)
Management Brief / Report (recommendation + next steps)