GenAI - Technical Challenge

Context

You work on a team building a GenAI-powered tool to help car mechanics. Mechanics enter free-text descriptions of vehicle issues, and your task is to **map these descriptions to the correct diagnostic codes**.

You are provided with:

1. **A dictionary of diagnostic codes** (dictionary.json)
2. **A small dataset of mechanic text inputs** with their corresponding diagnostic codes (sample_data.json).

`IMPORTANT: Training or finetuning a LLM is not expected and out of scope for this challenge.`

Task

1. Exploratory Understanding

- Inspect the diagnostic code dictionary (dictionary.json).
- Explore the mechanic notes dataset (sample_data.json).
- Identify challenges (e.g., spelling errors, abbreviations, informal language). DO NOT modify the given data set to address these challenges!

2. Baseline Approach (Zero-/Few-Shot Prompting)

- Using an LLM (OpenAI API, Hugging Face, or similar), create a **prompting approach** that:
  - Takes in a mechanic's note.
  - Suggests the most relevant diagnostic code(s).
- Document how you designed your prompt and why.

3. Improved Approach (Dictionary-Aware Retrieval)

- Implement a **retrieval-augmented method** where:
  - Mechanic note is embedded (sentence transformer or OpenAI embeddings).
  - You retrieve the **top-N diagnostic codes** from the dictionary.
- Then either:
  - Use an LLM to choose the best code from those candidates, **or**
  - Score the similarity and output the top match.

4. Testing

- Test both approaches (prompt-only vs retrieval-assisted) on the provided dataset.
- Compute simple metrics: accuracy, top-3 accuracy.
- Give 1–2 examples of failures and your thoughts on how to improve.

Deliverables

- A short Jupyter Notebook, Google Colab link, or Python script implementing the steps.
- A brief README (1–2 pages) describing **your approach**, **design decisions**, and **results/takeaways**.