AI Engineering Technical Assignment: Intelligent Dog Breed Assistant

Overview

Create an intelligent question-answering system that helps users find information about dog breeds. This project will assess your ability to build AI systems that can handle both natural language understanding and data analytics.

The Challenge

Design and implement a smart API endpoint that can answer questions about dog breeds using a comprehensive dataset. Your solution should demonstrate expertise in both natural language processing and data analysis by effectively handling two distinct types of queries:

1. Natural Language Understanding: Questions requiring semantic comprehension and contextual understanding

2. Data Analytics: Queries needing numerical analysis and precise data manipulation

Technical Requirements

User Interface

Create a simple, functional web interface for your system using any framework of your choice (e.g., Streamlit, Gradio, Flask+HTML, etc.). The UI should:

- Allow users to input questions about dog breeds
- Display the system's responses clearly
- Maintain a basic conversation history

- Have a clean, functional design

Note: The focus is on functionality rather than design aesthetics. Keep the UI simple and intuitive.

Containerization

Your solution must be containerized using Docker:

- Include a Dockerfile for building your application
- Provide docker-compose.yml if using multiple services
- Ensure all dependencies are properly managed
- Include clear instructions for building and running the container

Data

You will work with a rich dataset containing detailed information about dog breeds, including:

- Physical attributes (height, weight, lifespan)
- Behavioral characteristics and temperament
- Care requirements
- Training attributes
- Breed classifications
- Historical information
- Popularity metrics

Access the dataset here: [Dog Breeds Dataset] (https://www.kaggle.com/api/v1/datasets/download/mexwell/dog-breeds-dataset)

Core Implementation

API Endpoint Specification

Create a single endpoint with the following interface:

```
# Input Parameters:

{
   "user_id": string,  # Unique identifier for the user
   "query": string     # User's question about dog breeds
}

# Response:

{
"answer": string    # Generated response to the query
}
```

Required Capabilities

1. Natural Language Processing Pipeline

Your system should effectively handle queries such as:

- "I have young kids and limited time for grooming. Which breed would suit my family?"
- "What breeds are known for being both protective and good with families?"
- "I'm looking for a tall, graceful dog with a flowing coat and independent personality"

2. Data Analysis Pipeline

Your system should process analytical queries like:

- "List the 5 most popular breeds in the dataset"
- "Which breeds live the longest on average?"
- "Show me all large dogs (over 60cm) ordered by weight"

Evaluation Criteria

Your solution will be assessed on:

- **Code quality and organization**
- **Effectiveness of both NLP and analytical processing pipelines**
- **Accuracy and relevance of responses**
- UI functionality and usability
- Docker implementation and configuration
- Repository organization and documentation

Submission Guidelines

Create a GitHub repository containing:

1. Complete source code
2. Dockerfile and any related container configuration
3. Requirements/dependency files
4. Provide a comprehensive README.