ML QA Automation Engineer Task

*focused on key areas of ensuring quality and reliability:-*

As part of our technical assessment, we would like to understand how you would approach quality assurance for a project involving a machine learning-powered chatbot. The chatbot is designed to provide technical support for customers of a tech company specializing in consumer electronics. The key functionalities include troubleshooting common issues, providing step-by-step guides, and offering information on warranty and repair services. Please detail how you would address the following three critical areas:

1. **Data Quality Assurance**:  
   - Write a Python script to validate and preprocess a dataset of FAQ documents.

Instructions:  
     - Implement functions to clean and preprocess the text data (e.g., removing HTML tags, converting text to lowercase).  
     - Validate the integrity of the data by checking for missing values and ensuring consistent formatting.  
     - Write automated tests using pytest to verify the correctness of the preprocessing steps.

Deliverable: Submit the script along with the test cases and a sample dataset.2. Model

**Performance Testing**:  
   - Develop test cases to evaluate the performance of the machine learning models using relevant metrics.

   Instructions:  
     - Write tests to measure accuracy, precision, recall, and F1 score of the models.  
     - Implement cross-validation to ensure the models generalize well to unseen data.  
     - Validate the model outputs against a set of predefined queries and expected responses.  
   - Submit the test cases and a report on model performance metrics.3.

**API Endpoint Testing**:  
   - Develop an automated test suite for the `/query` endpoint in FastAPI that handles user queries.  
   Instructions:  
     - Write test cases for valid and invalid inputs to ensure the API returns the correct responses and error messages.  
     - Use mocking to simulate database interactions and verify that queries are correctly logged.  
     - Ensure the tests can be run as part of a CI/CD pipeline.  
   - Submit the test suite and instructions on how to run the tests.  
- Submit your code in a GitHub repository.  
- Include a README file with instructions on how to set up the project, run the tests, and any other relevant details.