Lab Activity: Building a Simple Chatbot with Local Models

Objective:

In this lab, you will build a basic chatbot in Python using both a local model (Ollama) and OpenAI's API. You will learn to:

1. Build a simple chatbot in local using Python's input and print functions.
2. Compare the ease of use and performance of a local model and OpenAI's API.

Part 1: Using a Local Model (Ollama)

Step 1: Install Dependencies

Before starting, ensure that you have the requests and json libraries installed. Run the following in your terminal:

```
$ pip install requests
```

Step 2: Install Ollama & Ensure it is Running

- Download Ollama and install it in your local machine:
  - https://ollama.com
- Run Ollama in your local machine using cmd:
  - ollama run <model-name>
- Make sure the Ollama server is running on your local machine.
  - http://localhost:11434

Step 3: Write the Python Code

Python code in chatbot.py file to creates a chatbot that interacts with a local model through the Ollama API. Use simple functions like print and input for the UI.

Step 4: Run the Program

```
python chatbot.py
```

Step 5: Enter some prompts, and see how the local model responds. Compare the performance and output with different inputs.

Part 2: Compare the Performance and Output

Now that you have two working chatbot systems:

1. **Run both systems**: Try various prompts on both the local and OpenAI models.
2. **Observe response times**: Does the local model respond faster or slower than OpenAI? How about accuracy and creativity in the answers?
3. **Quality of Responses**: Are there significant differences in the content of the responses?

Discussion Questions:

1. **Which model provided faster responses? Why do you think that is?**
2. **Which model gave more accurate or creative responses?**
3. **Based on your experience, how do you feel about integrating these models into a web application using something like Flask or Streamlit?**