

Setup Guide

Follow these steps to set up and run the RAG Pipeline Chatbot on your local machine.

1. Install Ollama

- Ollama is required to run the local LLM model.
- Download & Install Ollama (if not already downloaded) from Ollama Official Website : <https://ollama.com/download>
- Follow the installation instructions for your operating system.
- Pull the Model phi3:mini using the command : `"ollama pull phi3:mini"`
- Start Ollama in the Background
- Ollama should be running while using the chatbot. Usually, Ollama starts automatically after installation, but you can start it manually if needed using : `"ollama serve"`.

2. Install Tesseract OCR

- Tesseract is required for processing images and scanned PDFs.
- For Windows download the installer from Tesseract GitHub Releases.
- Install it and note the installation path (e.g. C:\Program Files\Tesseract-OCR\tesseract.exe)

3. Configure Tesseract Path

- On Windows, you must set the path to Tesseract inside preprocessor.py:
- Open [preprocessor.py](#), go to line 21 and set the path like this:
`pytesseract.pytesseract.tesseract_cmd = r"C:\Program Files\Tesseract-OCR\tesseract.exe"`
- (Optional) Add Tesseract to your system PATH environment variable: Press Win + S → Search for Environment Variables. Edit the Path variable under system variables. Add the folder path where tesseract.exe is installed.

4. Clone the Repository

- git clone [<https://github.com/RaisaTahasen/Chatbot-RAG>]
- cd RaisaTahasen/Chatbot-RAG

5. Install Python Dependencies

- Make sure you have Python 3.9+ installed.
- Run `"pip install -r requirements.txt"` in your terminal

6. Run the Application

- Start the Streamlit app: `"streamlit run app_streamlit.py"`
- Open the app in your browser: <http://localhost:8501>

7. Use the Chatbot

- Upload one or more files (.csv, .txt, .docx, .sqlite, .db, .pdf, .jpeg, .jpg, .png) via the sidebar.

- Wait for processing (indicator will turn green when ready).

8. Enter your question in the main panel.

- Click **Get Answer**. [Note: First query may take some time as the model loads.]