SETUP:
Introduction to RAG with Ollama

Contents

Source Material

The Jupyter Notebooks that we'll use for this session were created by Nishad Thalhath for a workshop session at SWIB 2024, and are available on GitHub:

https://github.com/nishad/llm-workshop-notebooks/tree/main

Requirements

From llm-workshop-notebooks/README.md

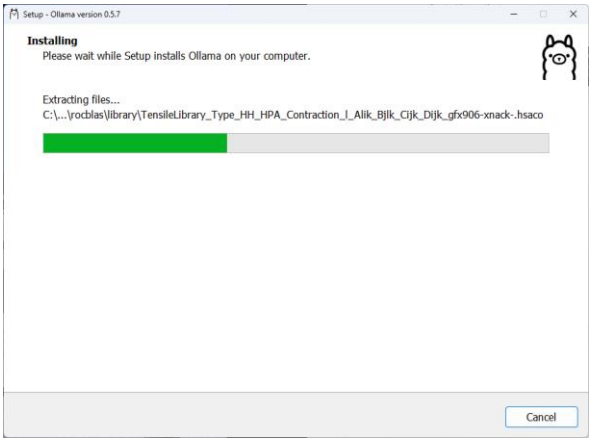
Participants are expected to have:

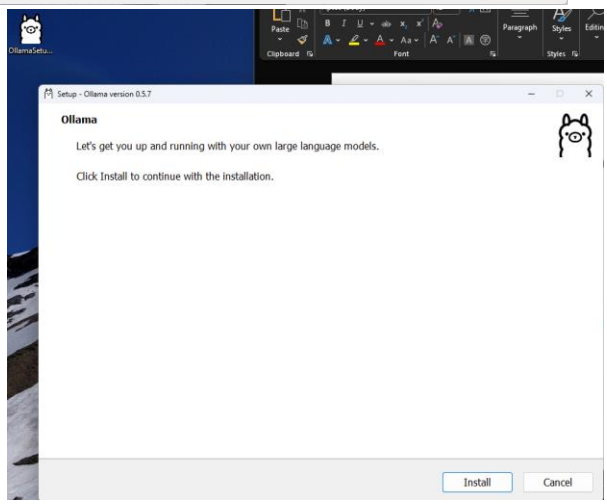
- A computer with at least 8GB RAM and 20GB free disk space.
- A recent version of Python installed with Jupyter Notebook support.
- Basic understanding of Python programming (for Part 2).
- A stable internet connection for downloading models and tools.

Windows setup

Download + install Ollama

https://ollama.com/download

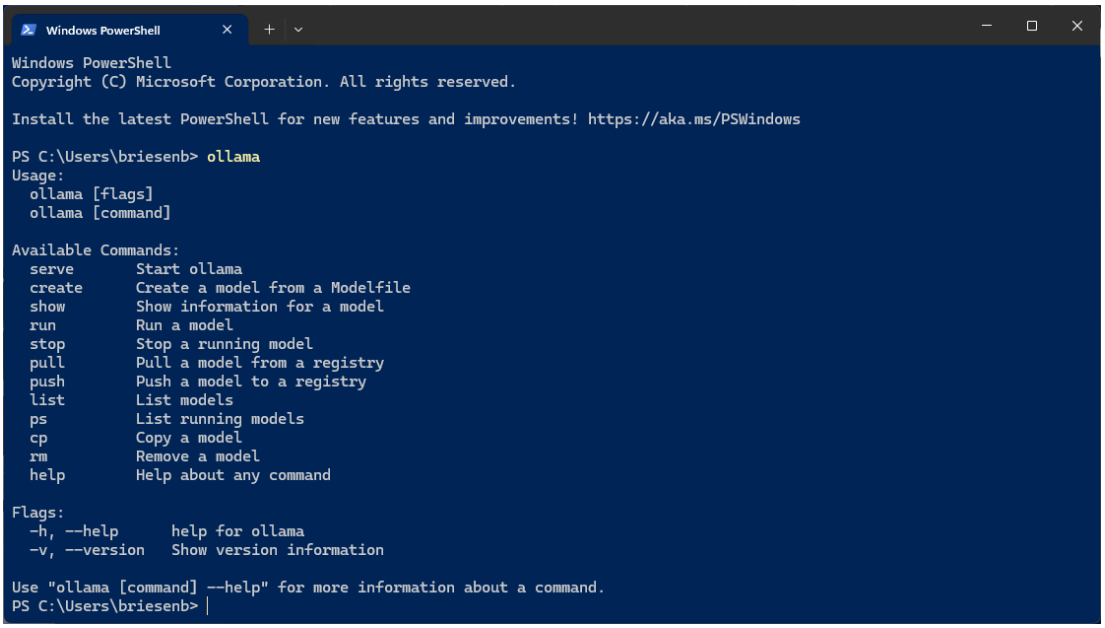


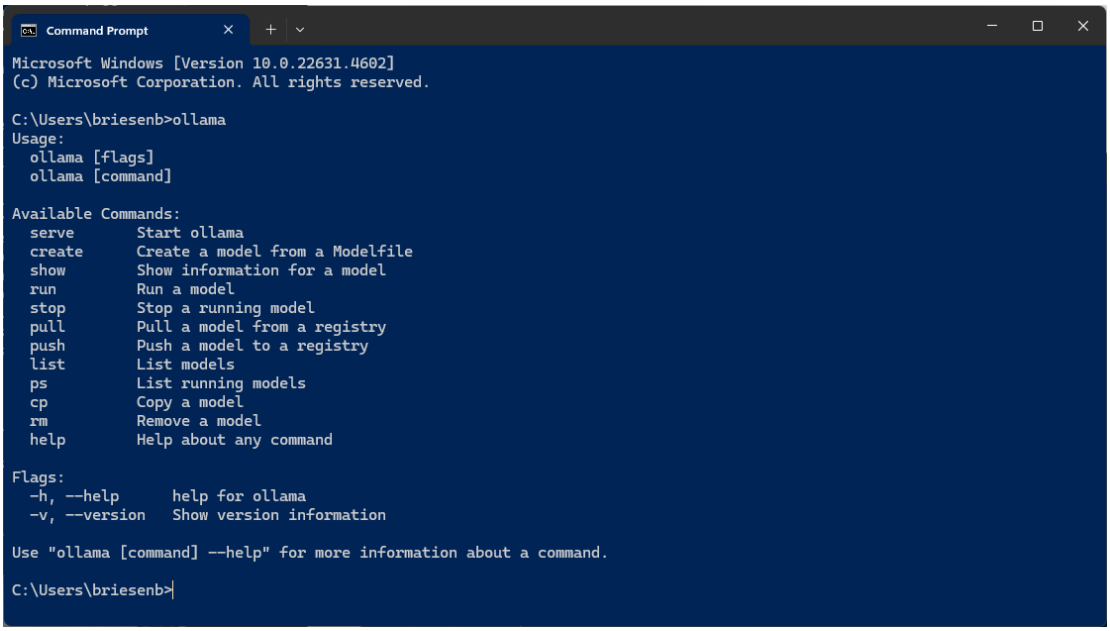


Confirm Ollama install

Open Powershell or Windows Terminal
(I'll use PowerShell for the rest of the Windows terminal examples.)

```
> ollama
# …help text appears in terminal
```

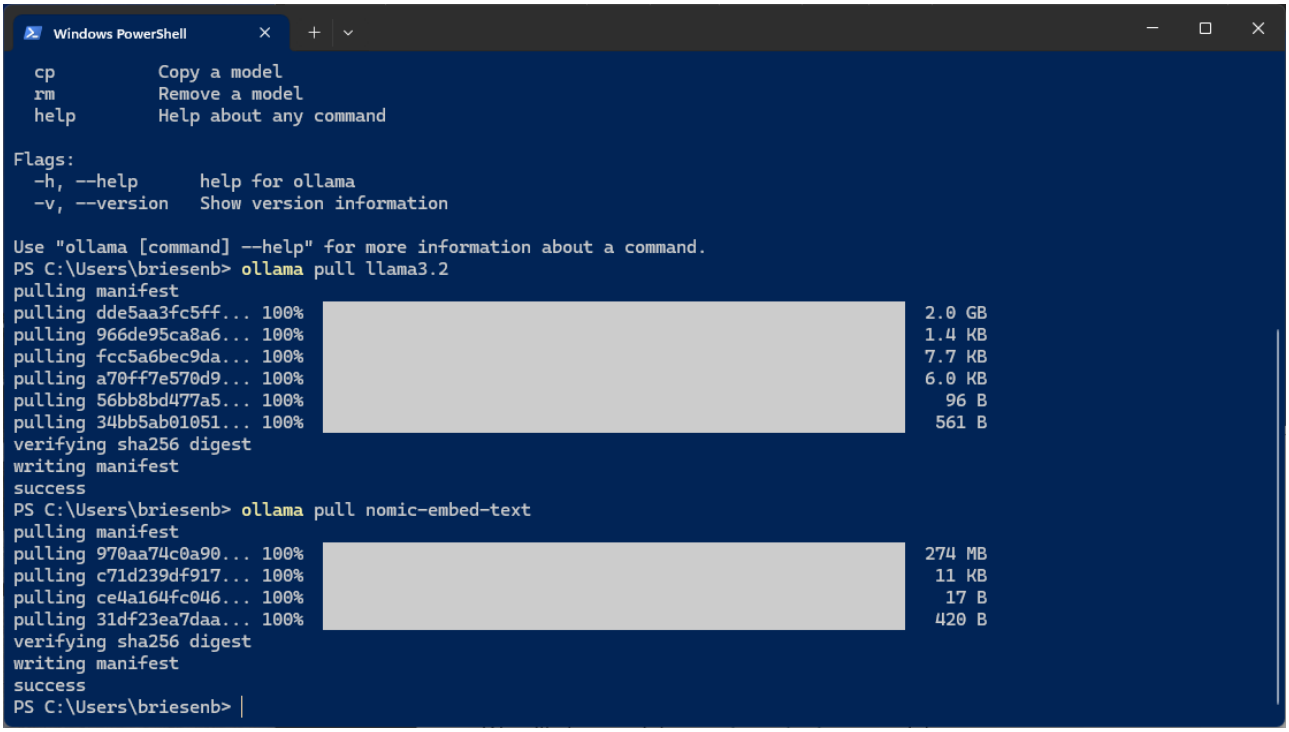




Get models needed for demo

See Ollama > Models
https://ollama.com/search
llama 3.3 is the “New state of the art 70B [parameter] model” but it is 43GB.
I’m installing llama3.2 > 3b, which is 2.0GB.
We will also need the nomic-embed-text model.

```
> ollama pull llama3.2
# …
> ollama pull nomic-embed-text
# …
```



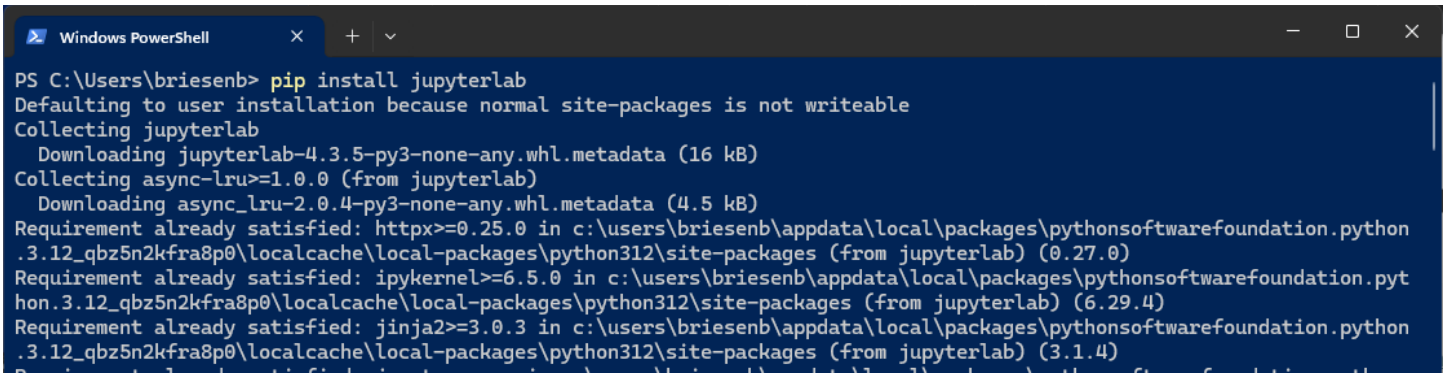
Prepare to run Python code and install Python packages

Python 3.12 can be installed from the Microsoft Store, or by downloading from Python.org
pip, a package installer for Python, may be automatically installed, or you may need to install from the command line. Further details are available in pip documentation > Installation

Prepare to work in Jupyter Notebook files (.ipynb)

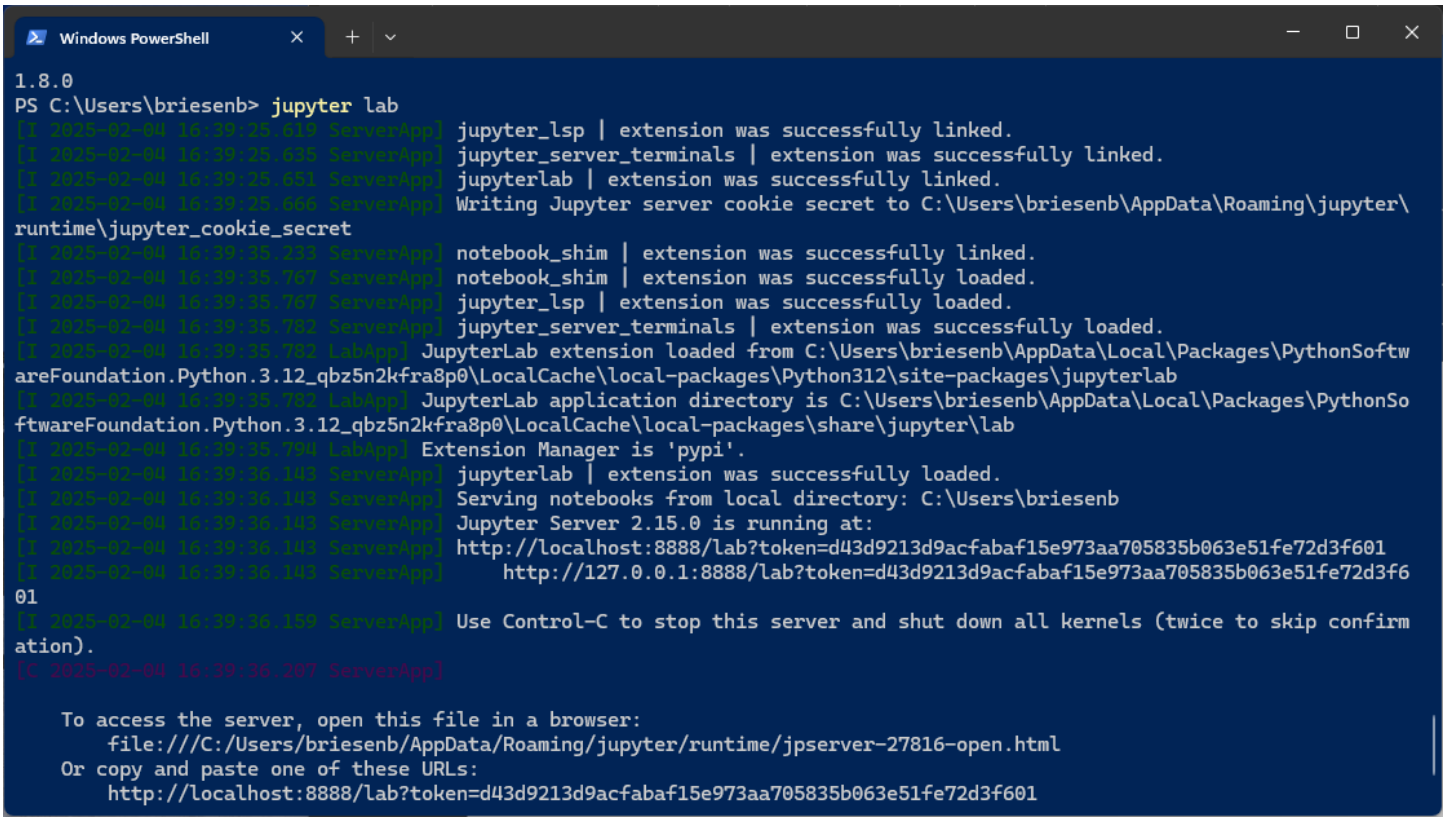
JupyterLab can be installed from the command line using pip (see additional details here)

```
> pip install jupyterlab
# …
```



Run Jupyter Lab

```
> jupyter lab
# … Jupyter lab will open in browser
```



Create notebook, install Ollama Python Library, confirm installed models

https://pypi.org/project/ollama/

See installation of Ollama models above