

SETUP:

Introduction to RAG with Ollama

Contents

Source Material	1
Requirements	1
Windows setup	2
Download + install Ollama https://ollama.com/download	2
Confirm Ollama install	3
Get models needed for demo	4
Prepare to run Python code and install Python packages	4
Prepare to work in Jupyter Notebook files (.ipynb)	5
Run Jupyter Lab	5
Create notebook, install Ollama Python Library, confirm installed models	5

Source Material

The Jupyter Notebooks that we'll use for this session were created by [Nishad Thalath](#) 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

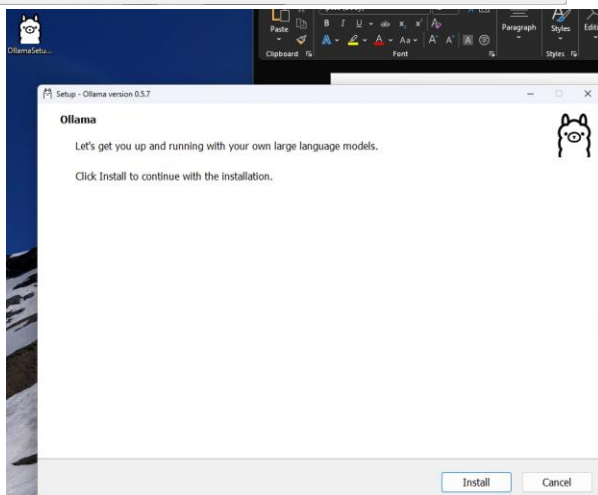
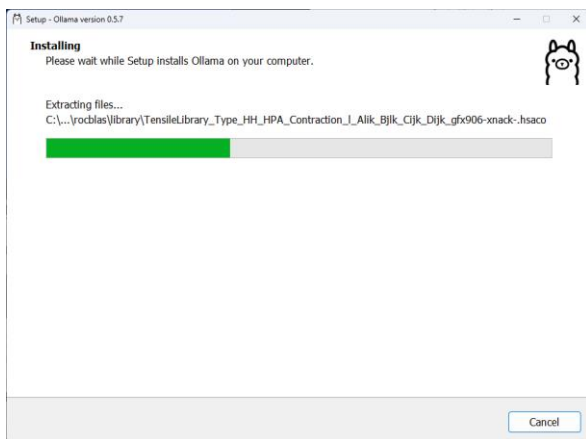
<https://ollama.com/download>

Download Ollama



Download for Windows

Requires Windows 10 or later



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

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\briesenb> ollama
Usage:
  ollama [flags]
  ollama [command]

Available Commands:
  serve      Start ollama
  create     Create a model from a Modelfile
  show       Show information for a model
  run        Run a model
  stop       Stop a running model
  pull       Pull a model from a registry
  push       Push a model to a registry
  list       List models
  ps         List running models
  cp         Copy a model
  rm         Remove a model
  help       Help about any command

Flags:
  -h, --help      help for ollama
  -v, --version    Show version information

Use "ollama [command] --help" for more information about a command.
PS C:\Users\briesenb> |
```

```
Command Prompt
Microsoft Windows [Version 10.0.22631.4602]
(c) Microsoft Corporation. All rights reserved.

C:\Users\briesenb>ollama
Usage:
  ollama [flags]
  ollama [command]

Available Commands:
  serve      Start ollama
  create     Create a model from a Modelfile
  show       Show information for a model
  run        Run a model
  stop       Stop a running model
  pull       Pull a model from a registry
  push       Push a model to a registry
  list       List models
  ps         List running models
  cp         Copy a model
  rm         Remove a model
  help       Help about any command

Flags:
  -h, --help      help for ollama
  -v, --version    Show version information

Use "ollama [command] --help" for more information about a command.
C:\Users\briesenb>
```

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.

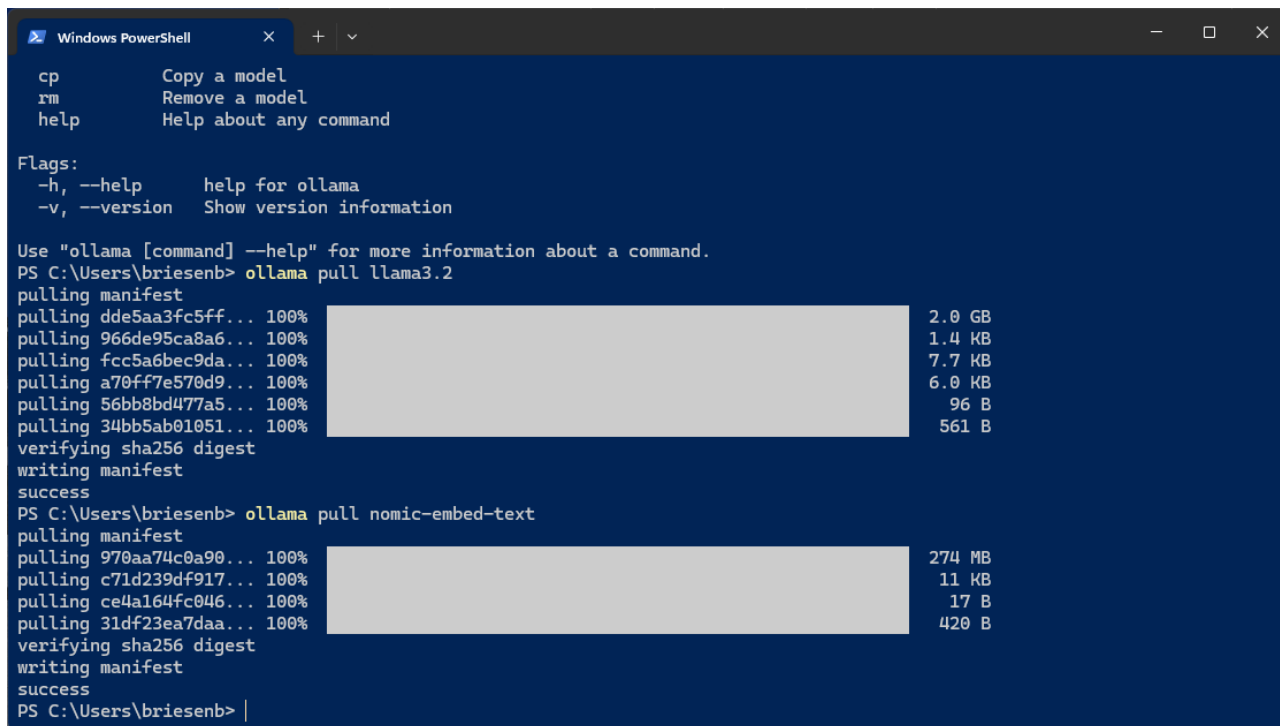
We will also need the [nomic-embed-text](#) model.

```
> ollama pull llama3.2
```

```
# ...
```

```
> ollama pull nomic-embed-text
```

```
# ...
```



```
Windows PowerShell
cp      Copy a model
rm      Remove a model
help    Help about any command

Flags:
-h, --help      help for ollama
-v, --version   Show version information

Use "ollama [command] --help" for more information about a command.
PS C:\Users\briesenb> ollama pull llama3.2
pulling manifest
pulling dde5aa3fc5ff... 100% 2.0 GB
pulling 966de95ca8a6... 100% 1.4 KB
pulling fcc5a6bec9da... 100% 7.7 KB
pulling a70ff7e570d9... 100% 6.0 KB
pulling 56bb8bd477a5... 100% 96 B
pulling 34bb5ab01051... 100% 561 B
verifying sha256 digest
writing manifest
success
PS C:\Users\briesenb> ollama pull nomic-embed-text
pulling manifest
pulling 970aa74c0a90... 100% 274 MB
pulling c71d239df917... 100% 11 KB
pulling ce4a164fc046... 100% 17 B
pulling 31df23ea7daa... 100% 420 B
verifying sha256 digest
writing manifest
success
PS C:\Users\briesenb> |
```

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](#)

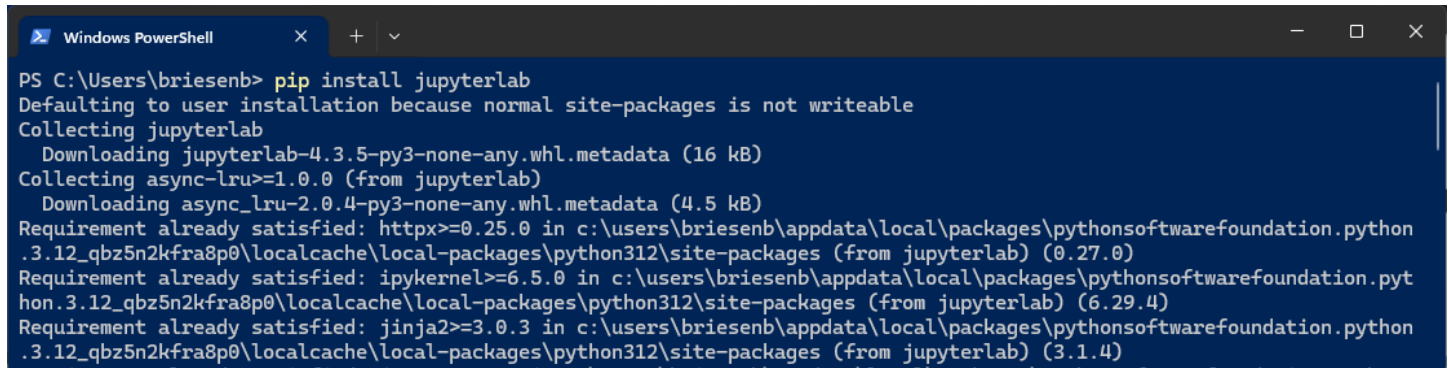
[pip](#), a package installer for Python, may be automatically installed, or you may need to install from the command line. Further details are available in pip documentation > [Installation](#)

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
```

```
# ...
```

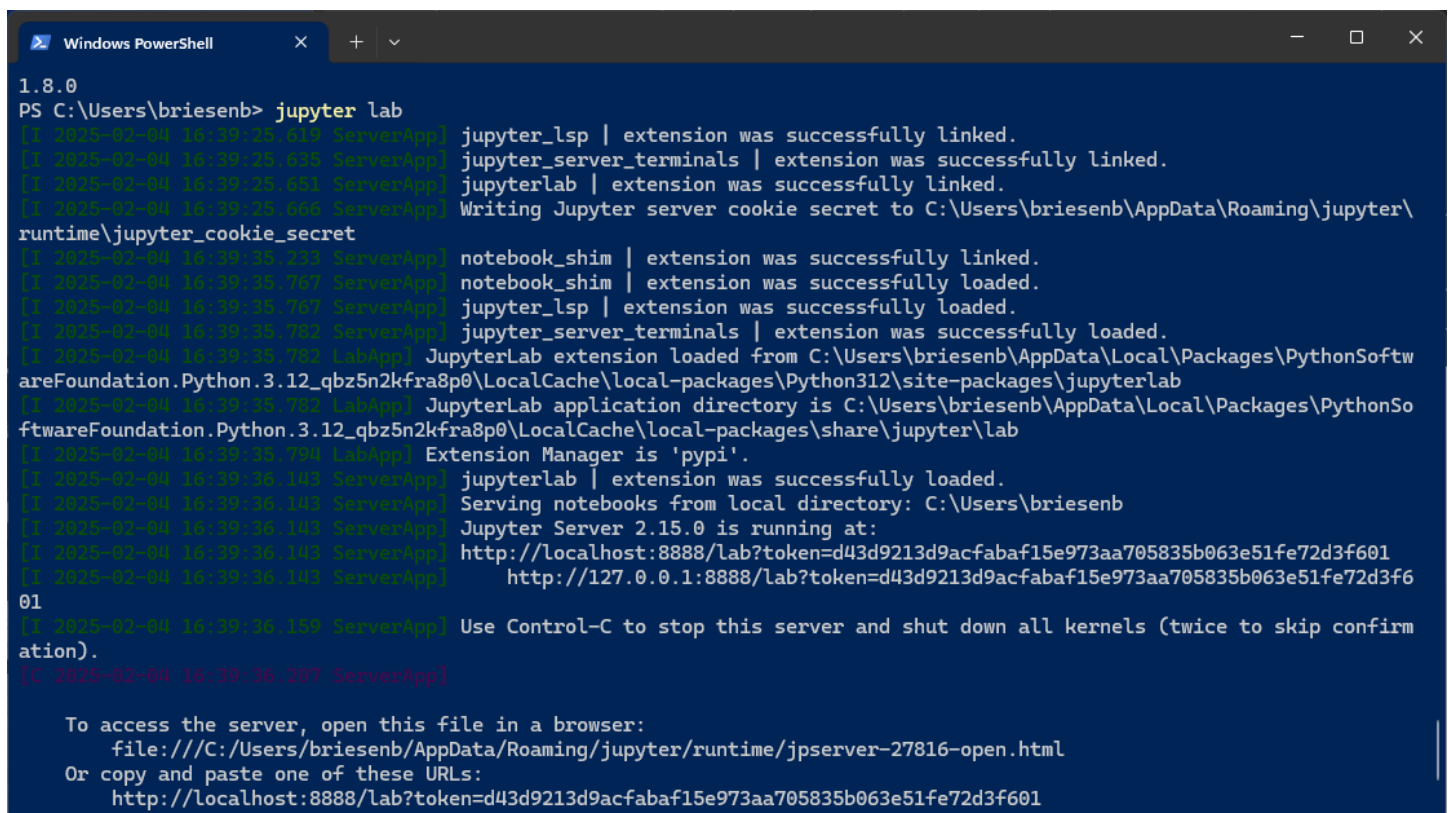


```
PS C:\Users\briesenb> pip install jupyterlab
Defaulting to user installation because normal site-packages is not writeable
Collecting jupyterlab
  Downloading jupyterlab-4.3.5-py3-none-any.whl.metadata (16 kB)
Collecting async-lru>=1.0.0 (from jupyterlab)
  Downloading async_lru-2.0.4-py3-none-any.whl.metadata (4.5 kB)
Requirement already satisfied: httpx>=0.25.0 in c:\users\briesenb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from jupyterlab) (0.27.0)
Requirement already satisfied: ipykernel>=6.5.0 in c:\users\briesenb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from jupyterlab) (6.29.4)
Requirement already satisfied: jinja2>=3.0.3 in c:\users\briesenb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from jupyterlab) (3.1.4)
```

Run Jupyter Lab

```
> jupyter lab
```

```
# ... Jupyter lab will open in browser
```



```
1.8.0
PS C:\Users\briesenb> jupyter lab
[1 2025-02-04 16:39:25.619 ServerApp] jupyter_lsp | extension was successfully linked.
[1 2025-02-04 16:39:25.635 ServerApp] jupyter_server_terminals | extension was successfully linked.
[1 2025-02-04 16:39:25.651 ServerApp] jupyterlab | extension was successfully linked.
[1 2025-02-04 16:39:25.666 ServerApp] Writing Jupyter server cookie secret to C:\Users\briesenb\AppData\Roaming\jupyter\runtime\jupyter_cookie_secret
[1 2025-02-04 16:39:35.233 ServerApp] notebook_shim | extension was successfully linked.
[1 2025-02-04 16:39:35.787 ServerApp] notebook_shim | extension was successfully loaded.
[1 2025-02-04 16:39:35.787 ServerApp] jupyter_lsp | extension was successfully loaded.
[1 2025-02-04 16:39:35.782 ServerApp] jupyter_server_terminals | extension was successfully loaded.
[1 2025-02-04 16:39:35.792 LabApp] JupyterLab extension loaded from C:\Users\briesenb\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.12_qbz5n2kfra8p0\LocalCache\local-packages\Python312\site-packages\jupyterlab
[1 2025-02-04 16:39:35.792 LabApp] JupyterLab application directory is C:\Users\briesenb\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.12_qbz5n2kfra8p0\LocalCache\local-packages\share\jupyter\lab
[1 2025-02-04 16:39:35.794 LabApp] Extension Manager is 'pypi'.
[1 2025-02-04 16:39:36.143 ServerApp] jupyterlab | extension was successfully loaded.
[1 2025-02-04 16:39:36.143 ServerApp] Serving notebooks from local directory: C:\Users\briesenb
[1 2025-02-04 16:39:36.143 ServerApp] Jupyter Server 2.15.0 is running at:
[1 2025-02-04 16:39:36.143 ServerApp] http://localhost:8888/Lab?token=d43d9213d9acfabaf15e973aa705835b063e51fe72d3f601
[1 2025-02-04 16:39:36.143 ServerApp] http://127.0.0.1:8888/Lab?token=d43d9213d9acfabaf15e973aa705835b063e51fe72d3f601
[1 2025-02-04 16:39:36.189 ServerApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 2025-02-04 16:39:36.207 ServerApp]

To access the server, open this file in a browser:
file:///C:/Users/briesenb/AppData/Roaming/jupyter/runtime/jpserver-27816-open.html
Or copy and paste one of these URLs:
http://localhost:8888/Lab?token=d43d9213d9acfabaf15e973aa705835b063e51fe72d3f601
```

Create notebook, install Ollama Python Library, confirm installed models

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

See [installation of Ollama models above](#)

