

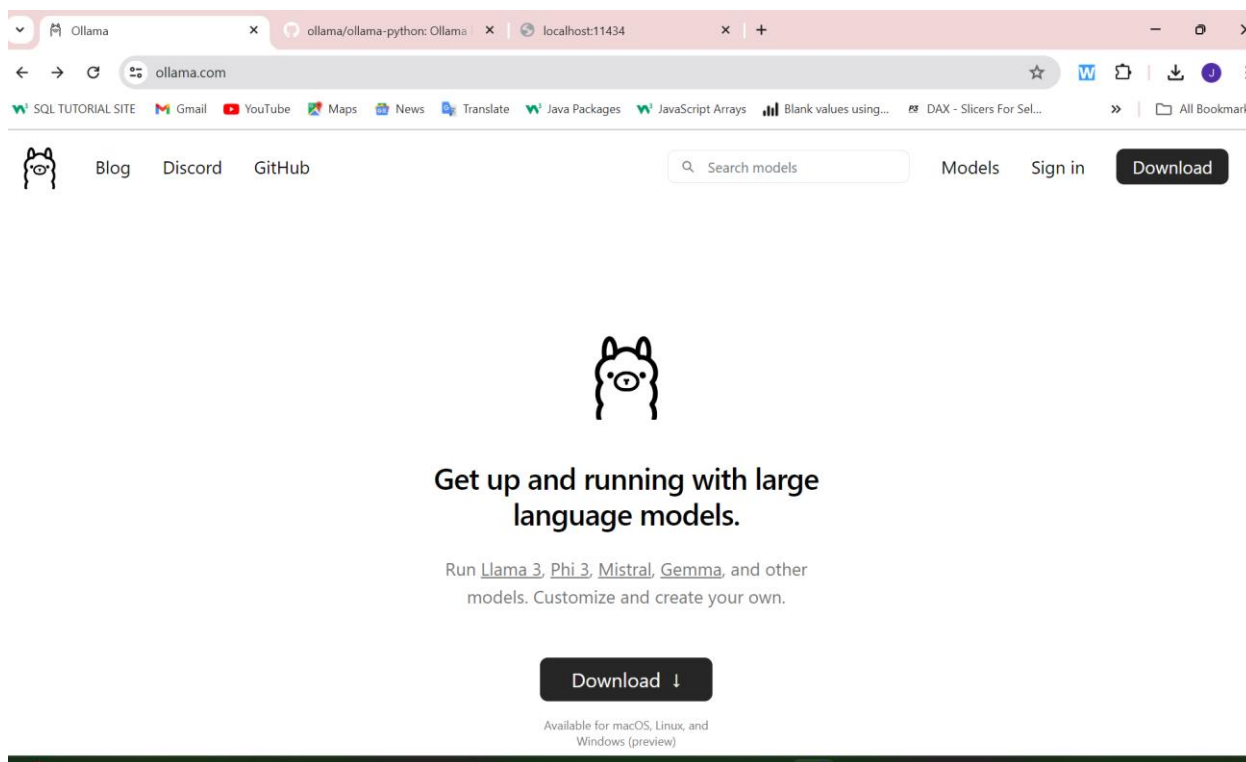
# Ollama-Local-LLM

Getting started with Ollama and self hosting Large Language Models for local AI solutions.

Welcome to our Ollama Local Setup Tutorial! I'm Juilee, and I'll be guiding you through installing and configuring Ollama on your own machines. Let's dive in and unlock the full potential of this powerful data processing tool together!

## Setup steps

1] Go to ollama website : <https://ollama.com/>



2] Download the Installer on Windows

<https://ollama.com/blog/windows-preview>

once downloaded double-click the installer, **OllamaSetup.exe**

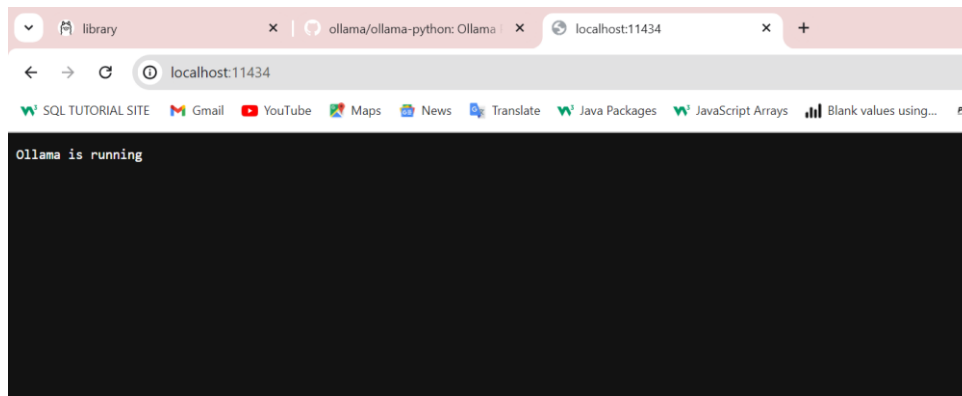


3] Once installed, go to terminal and view downloaded or manually created models, in terminal: **ollama list**

4] You can download any LLM of your choice: **ollama run tinyllama**

5] Once LLM is downloaded it will automatically be served on '<http://localhost:11434>' (Ollama's API automatically runs in the background, serving on <http://localhost:11434>. Tools and applications can connect to it without any additional setup.)

6] To serve LLM manually, type in terminal: **ollama serve**



```
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\juile> ollama list
NAME      ID          SIZE    MODIFIED
PS C:\Users\juile> ollama run tinyllama
pulling manifest
pulling af0ddbdaaa26... 100% 637 MB
pulling c8472cd9daed... 100% 70 B
pulling fa956ab37b8c... 100% 31 B
pulling 6331358be52a... 100% 98 B
verifying sha256 digest
writing manifest
removing any unused layers
success
```

7] After successfully downloading your model is running. Ask any question . It will generate response below.

```
removing any unused layers
success
>>> what is prompt engineering?
Prompt Engineering is an automated writing system that generates academic papers, essays, and other types of
providing specific instructions or prompts to the system. The output generated by Prompt Engineering is typically
high-quality, academically sound, and free from errors. It's a powerful tool for academic researchers, students,
and professionals who need assistance with writing in various fields such as business, science, law, and more.

Prompt Engineering's key features include:

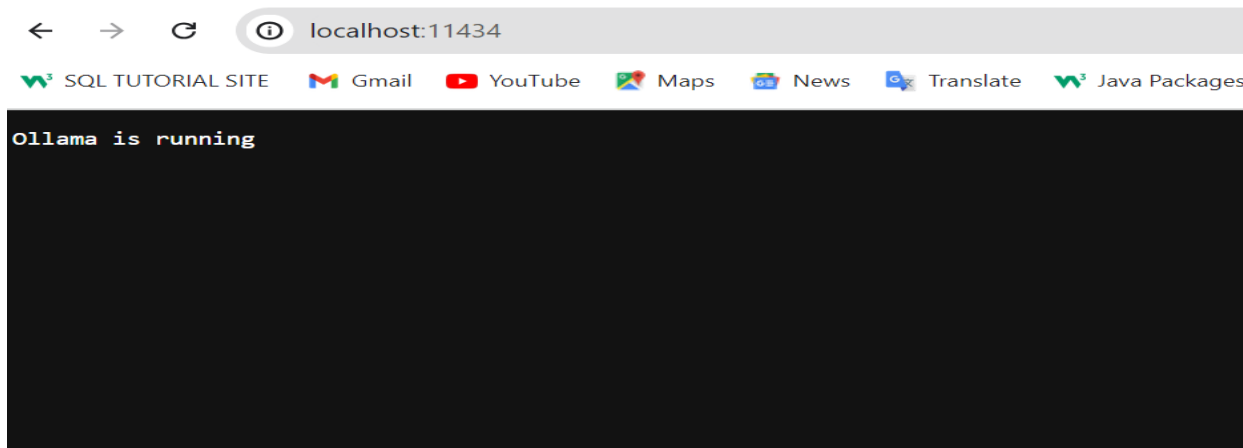
1. Robust NLI capabilities: Prompt Engineering offers a wide range of prompts, including natural language
questions, free-text input, and structured prompts such as tables or matrices.

2. Different types of content available: The system can generate academic essays, research reports, case studies,
case analyses, reports, technical reports, and more.

3. Automatic formatting and citation style support: Prompt Engineering automatically formats the generated content
according to academic conventions, including referencing styles such as APA, MLA, Harvard, Chicago, and others.

4. Customizable user interface: The system can be customized with your preferences, allowing you to choose between
different templates or types of content.

5. Integration with other tools: Prompt Engineering integrates seamlessly with other tools such as Microsoft
Office, LaTeX, and more. It also supports a range of file formats, including PDFs, Word documents, and Excel
spreadsheets.
```

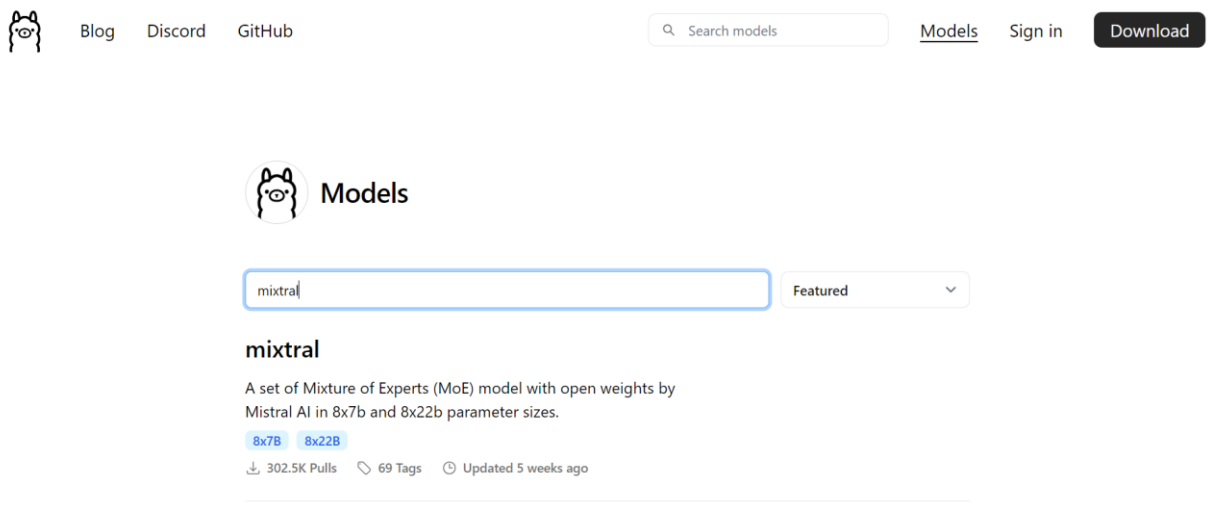


8] To exit the model click **ctrl+d**

Ollama supports numerous models from <https://ollama.com/library>

9] To view downloaded or manually created models, in terminal: **ollama list**

10] To download different models. Check the website. Click on Models. You can search the model of your choice.

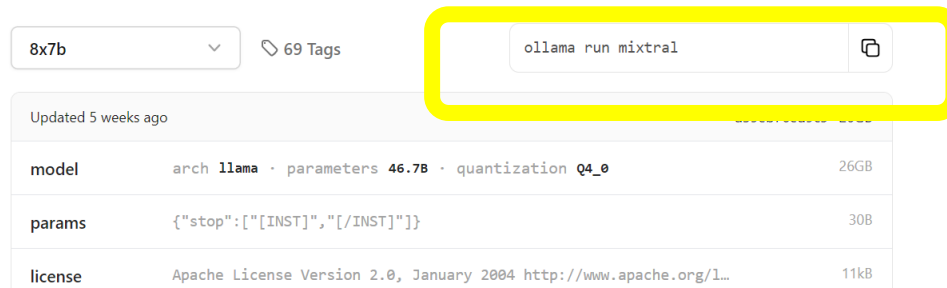


11] Copy and paste the command in the terminal.

## mixtral

A set of Mixture of Experts (MoE) model with open weights by Mistral AI in 8x7b and 8x22b parameter sizes.

[8x7B](#) [8x22B](#)  
302.5K Pulls Updated 5 weeks ago



12] The new model will get downloaded. Once done you can pass any prompt

### 13] To remove a model: `ollama rm mixtral`

Read more about usage with Ollama Python Library (<https://github.com/ollama/ollama-python>)

README.md

2 weeks ago

poetry.lock

Merge pull request #178 from ollama/dependabot/pip/pyt...

2 weeks ago

pyproject.toml

Bump ruff from 0.3.4 to 0.4.1

2 months ago

requirements.txt

update dependencies

3 months ago

README

MIT license

🔗 Ollama Python Library

The Ollama Python library provides the easiest way to integrate Python 3.8+ projects with [Ollama](#).

### Prerequisites

You need to have a local ollama server running to be able to continue. To do this:

- Download: <https://ollama.com/>
- Run an LLM: <https://ollama.com/library>
  - Example: `ollama run llama2`
  - Example: `ollama run llama2:70b`

Then:

Report repository















Releases 12

v0.2.0 Latest

on May 10

+ 11 releases

Contributors 17

+ 3 contributors

Deployments 16

release 2 weeks ago

+ 15 deployments

Languages

Python 100.0%