# Llama3

#### Llama3

Model size
Performance
Pull Llama3
Use Llama 3
Run Llama 3
Start a conversation
End the conversation
References

Demo Environment

**Development board**: Jetson Orin series motherboard

**SSD**: 128G

**Tutorial application scope**: Whether the motherboard can run is related to the available memory of the system. The user's own environment and the programs running in the background may cause the model to fail to run

| Motherboard model    | Ollama | Open WebUI |
|----------------------|--------|------------|
| Jetson Orin NX 16GB  | V      | √          |
| Jetson Orin NX 8GB   | V      | ×          |
| Jetson Orin Nano 8GB | V      | ×          |
| Jetson Orin Nano 4GB | ×      | ×          |

Meta Llama3 is a family of advanced open source Large Language Models (LLMs) developed by the Meta Al division.

### **Model size**

| Model  | Parameters |
|--------|------------|
| Llama3 | 8B         |
| Llama3 | 70B        |

### **Performance**

#### Meta Llama 3 Instruct model performance

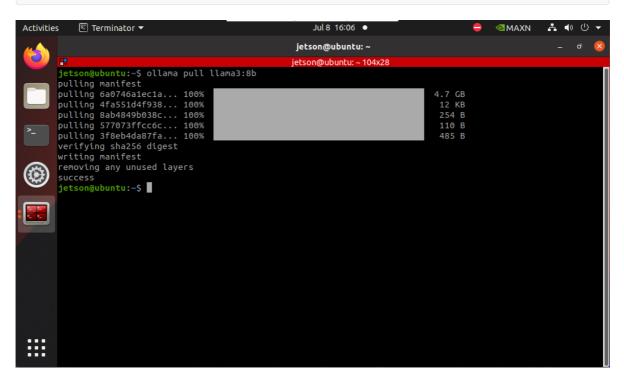
|                       | Meta<br>Llama 3<br>8B | Gemma<br>7B - It<br><sup>Measured</sup> | Mistral<br>7B Instruct<br>Measured |
|-----------------------|-----------------------|---|------------------------------------|
| MMLU<br>5-shot        | 68.4                  | 53.3                                    | 58.4                               |
| GPQA<br>0-shot        | 34.2                  | 21.4                                    | 26.3                               |
| HumanEval<br>0-shot   | 62.2                  | 30.5                                    | 36.6                               |
| GSM-8K<br>8-shot, CoT | 79.6                  | 30.6                                    | 39.9                               |
| MATH<br>4-shot, CoT   | 30.0                  | 12.2                                    | 11.0                               |

|                              | Meta<br>Llama 3<br>70B | Gemini<br>Pro 1.5<br>Published | Claude 3<br>Sonnet<br>Published |  |
|------------------------------|------------------------|--------------------------------|---------------------------------|--|
| MMLU<br>5-shot               | 82.0                   | 81.9                           | 79.0                            |  |
| GPQA<br>0-shot               | 39.5                   | <b>41.5</b><br>CoT             | <b>38.5</b><br>CoT              |  |
| HumanEval<br>0-shot          | 81.7                   | 71.9                           | 73.0                            |  |
| <b>GSM-8K</b><br>8-shot, CoT | 93.0                   | <b>91.7</b><br>11-shot         | <b>92.3</b><br>0-shot           |  |
| MATH<br>4-shot, CoT          | 50.4                   | <b>58.5</b><br>Minerva prompt  | 40.5                            |  |

## **Pull Llama3**

Using the pull command will automatically pull the model from the Ollama model library:

ollama pull llama3:8b



## **Use Llama 3**

### **Run Llama 3**

If the system does not have a running model, the system will automatically pull Llama3 8B model and run:

ollama run llama3:8b

#### Start a conversation

Why is the sky blue?

The time to reply to the question depends on the hardware configuration, please be patient!

```
Jul 8 16:14 •
                                                                                                                                                                                               MAXN

    Terminator ▼

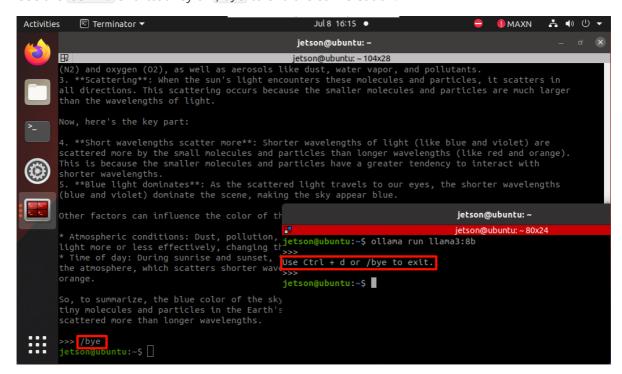
                                                                                                              jetson@ubuntu: ~
             jetson@ubuntu:~$ ollama pull llama3:8b
            pulling manifest
pulling 6a0746a1ec1a... 100%
pulling 4fa551d4f938... 100%
pulling 8ab4849b038c... 100%
                                                                                                                                                                      4.7 GB
                                                                                                                                                                        12 KB
                                                                                                                                                                        254 B
            pulling 577073ffcc6c... 100%
pulling 3f8eb4da87fa... 100%
verifying sha256 digest
writing manifest
                                                                                                                                                                         110 B
                                                                                                                                                                        485 R
            removing any unused layers
success
jetson@ubuntu:~$ ollama run llama3:8b
            >>> Why is the sky blue? What a great question!
             The sky appears blue because of a phenomenon called Rayleigh scattering, which is named after the
             British physicist Lord Rayleigh who first described it in the late 19th century.
            Here's what happens:
            1. **Light from the sun**: When sunlight enters Earth's atmosphere, it contains all the colors of the visible spectrum (red, orange, yellow, green, blue, indigo, and violet).

2. **Molecules and particles**: The atmosphere is made up of tiny molecules of gases like nitrogen (N2) and oxygen (O2), as well as aerosols like dust, water vapor, and pollutants.

3. **Scattering**: When the sun's light encounters these molecules and particles, it scatters in all directions. This scattering occurs because the smaller molecules and particles are much larger than the wavelengths of light.
             than the wavelengths of light.
```

#### **End the conversation**

Use the Ctrl+d shortcut key or /bye to end the conversation!



### References

Ollama

Official website: <a href="https://ollama.com/">https://ollama.com/</a>

GitHub: <a href="https://github.com/ollama/ollama">https://github.com/ollama/ollama</a>

## Llama 3

GitHub: <a href="https://github.com/meta-llama/llama3">https://github.com/meta-llama/llama3</a>

Ollama corresponding model: <a href="https://ollama.com/library/llama3">https://ollama.com/library/llama3</a>