WizardLM-2

WizardLM-2
- 1. Model scale
- 2. Pull WizardLM-2
- 3. Use WizardLM-2
  - 3.1. Run WizardLM-2
  - 3.2. Have a conversation
  - 3.3. 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	Run directly with Ollama	Run with Open WebUI
Jetson Orin NX 16GB	√	√
Jetson Orin NX 8GB	√	√
Jetson Orin Nano 8GB	√	√
Jetson Orin Nano 4GB	×	×

WizardLM-2 is Microsoft's latest advanced large-scale language model, which improves performance in complex dialogue, multi-language, reasoning, and intelligent agent capabilities.

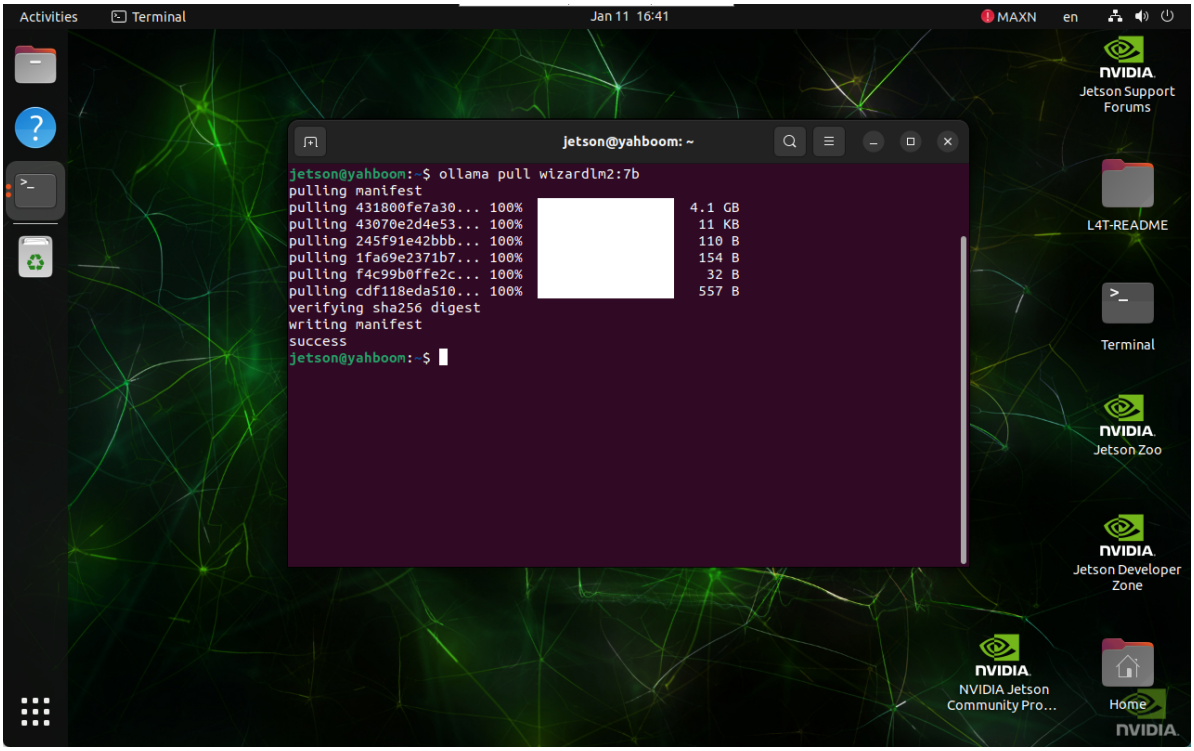
1. Model scale

Model	Parameters
WizardLM-2	7B
WizardLM-2	70B
WizardLM-2	8x22B

2. Pull WizardLM-2

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

```
ollama pull wizardlm2:7b
```



3. Use WizardLM-2

3.1. Run WizardLM-2

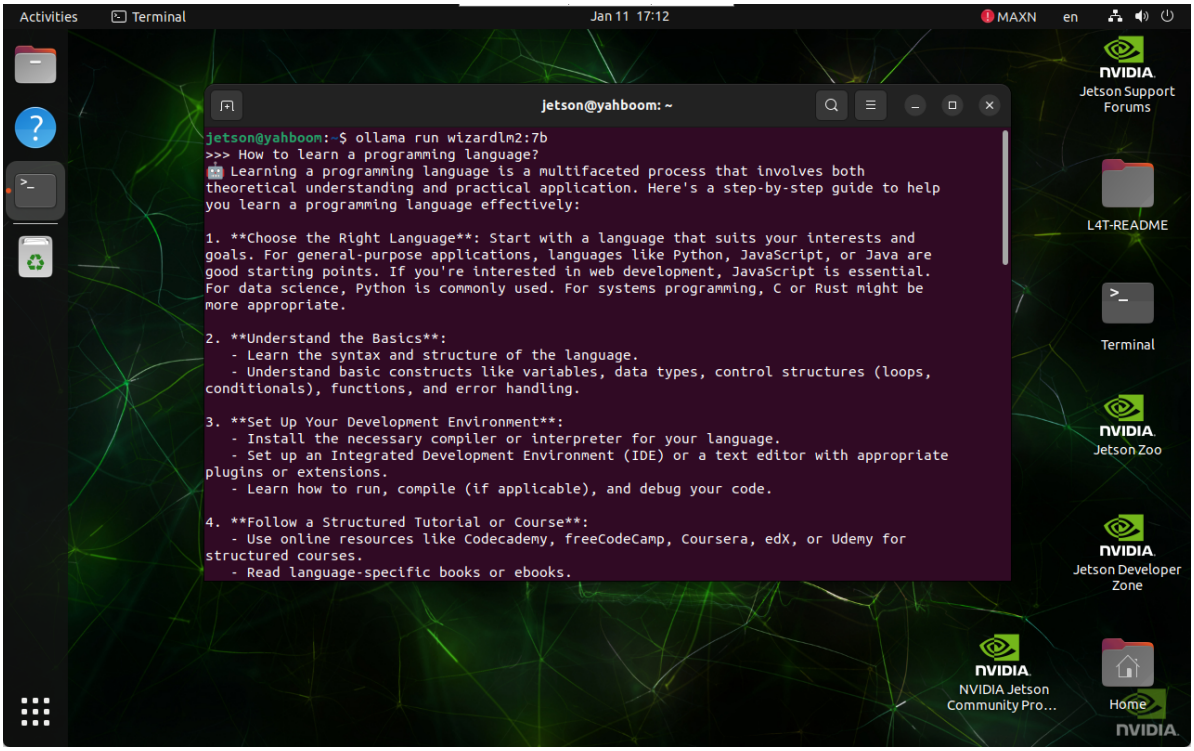
If the system does not have a running model, the system will automatically pull the WizardLM-2 7B model and run it:

```
ollama run wizardlm2:7b
```

3.2. Have a conversation

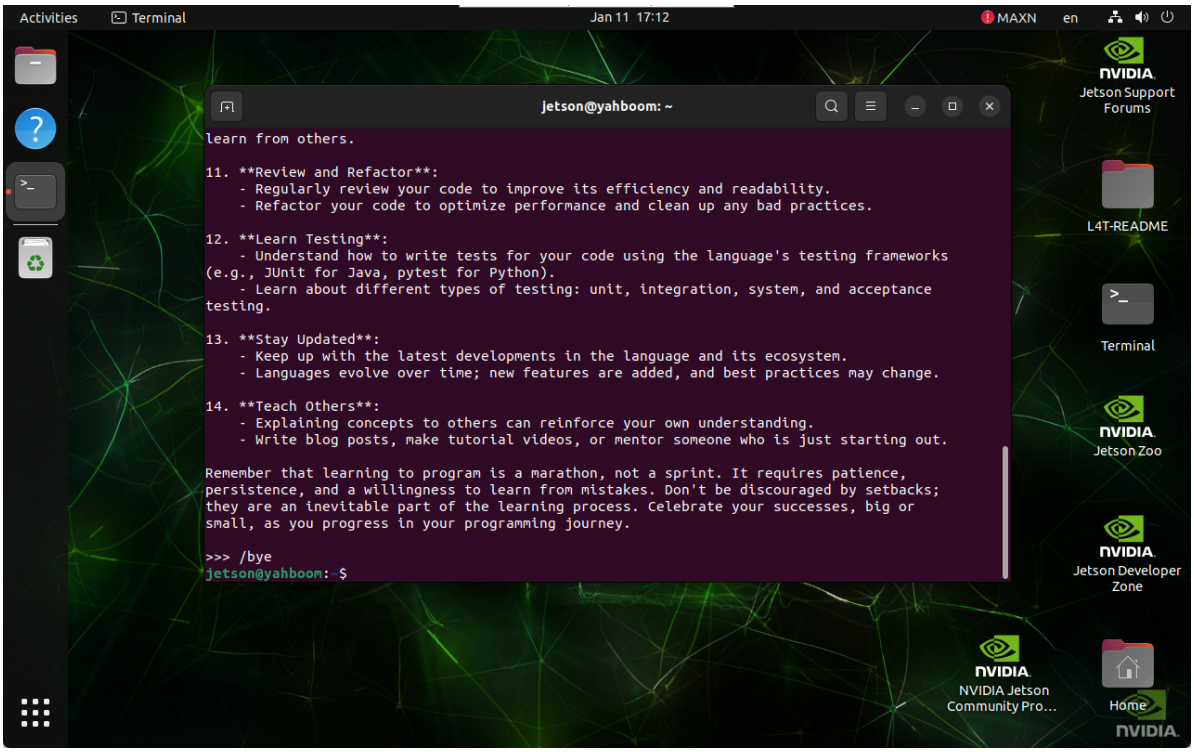
```
How to learn a programming language?
```

The time to reply to the question is related to the hardware configuration, please be patient!



3.3. End the conversation

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



References

> **Ollama**

Official website: https://ollama.com/

GitHub: https://github.com/ollama/ollama

> **WizardLM-2**

Ollama corresponding model: https://ollama.com/library/wizardlm2