DeepSeek Coder

> **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	√ (need to run the small parameter version)	√ (need to run the small parameter version)

DeepSeek Coder is an open source large language model (LLM) designed by DeepSeek to understand and generate code.

1. Model scale

Model	Parameters
DeepSeek Coder	1.3B
DeepSeek Coder	6.7B
DeepSeek Coder	33B

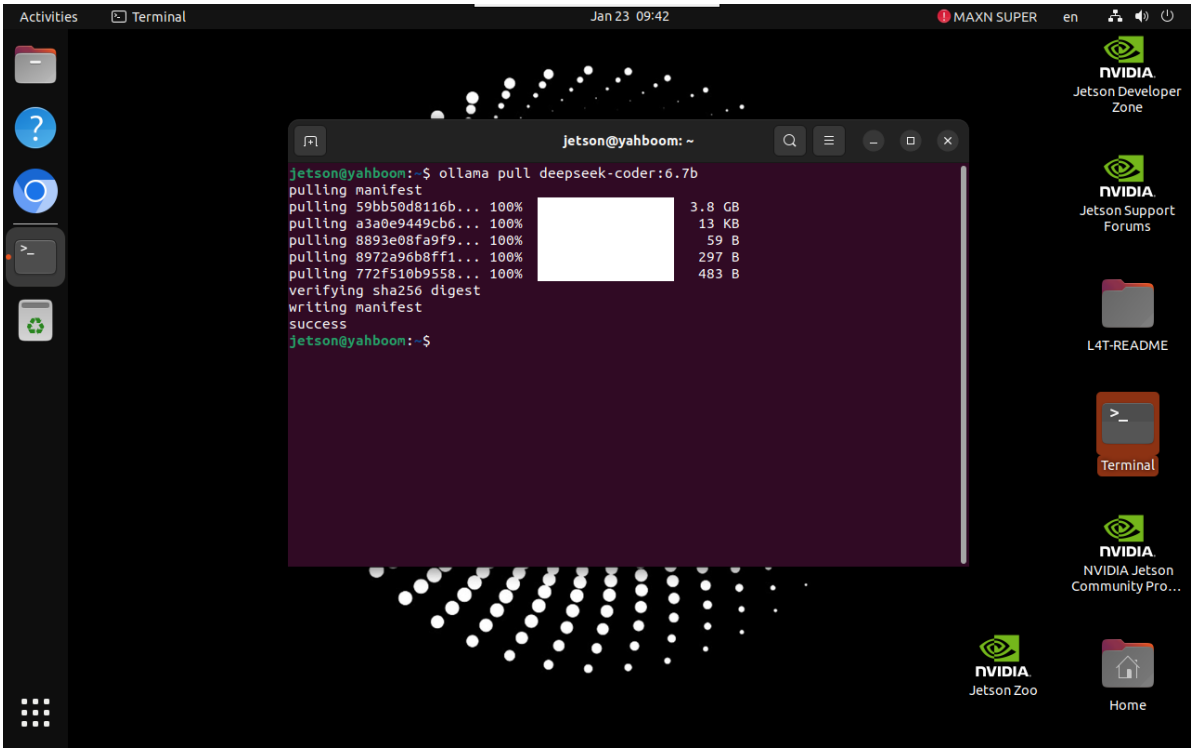
2. Pull DeepSeek Coder

Use the pull command to automatically pull the model of the Ollama model library:

```
ollama pull deepseek-coder:6.7b
```

Model with small parameters: motherboards with 8G memory or less can run this

```
ollama pull deepseek-coder:1.3b
```



3. Use DeepSeek Coder

3.1. Run DeepSeek Coder

If the system does not have a running model, the system will automatically pull the DeepSeek Coder 6.7B model and run it:

```
ollama run deepseek-coder:6.7b
```

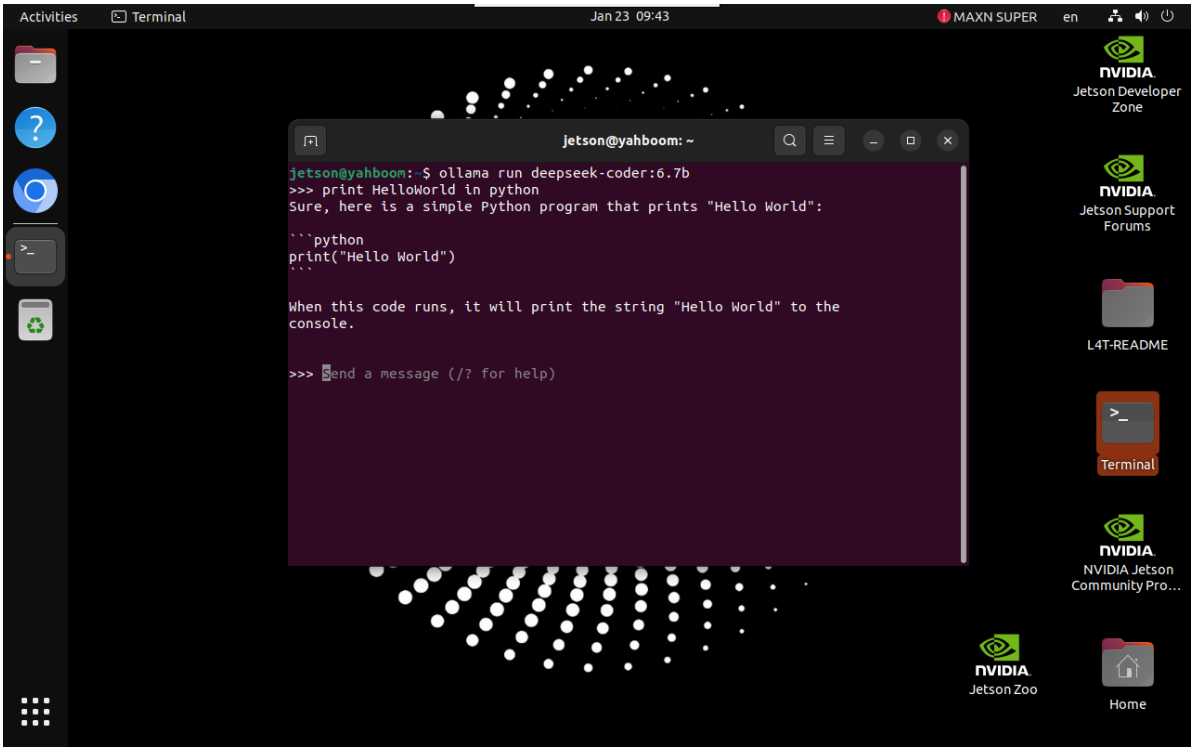
Model with small parameter version: motherboards with 8G memory or less can run this

```
ollama run deepseek-coder:1.3b
```

3.2. Have a conversation

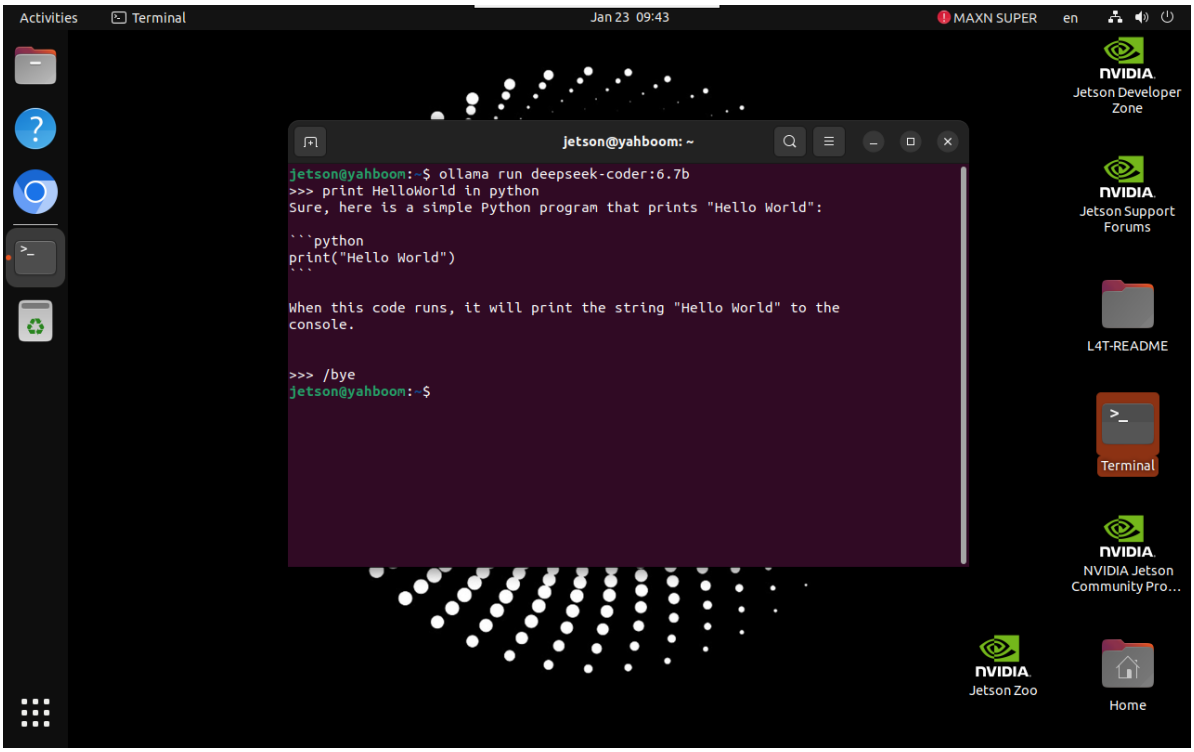
```
print HelloWorld in python
```

The time to reply to the question depends on 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

> **DeepSeek Coder**

Ollama corresponding model: https://ollama.com/library/deepseek-coder

GitHub: https://github.com/deepseek-ai/DeepSeek-Coder