

DeepSeek Coder

DeepSeek Coder

- Model size
- Pull DeepSeek Coder
- Use DeepSeek Coder
 - Run DeepSeek Coder
 - 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	√	√
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.

Model size

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

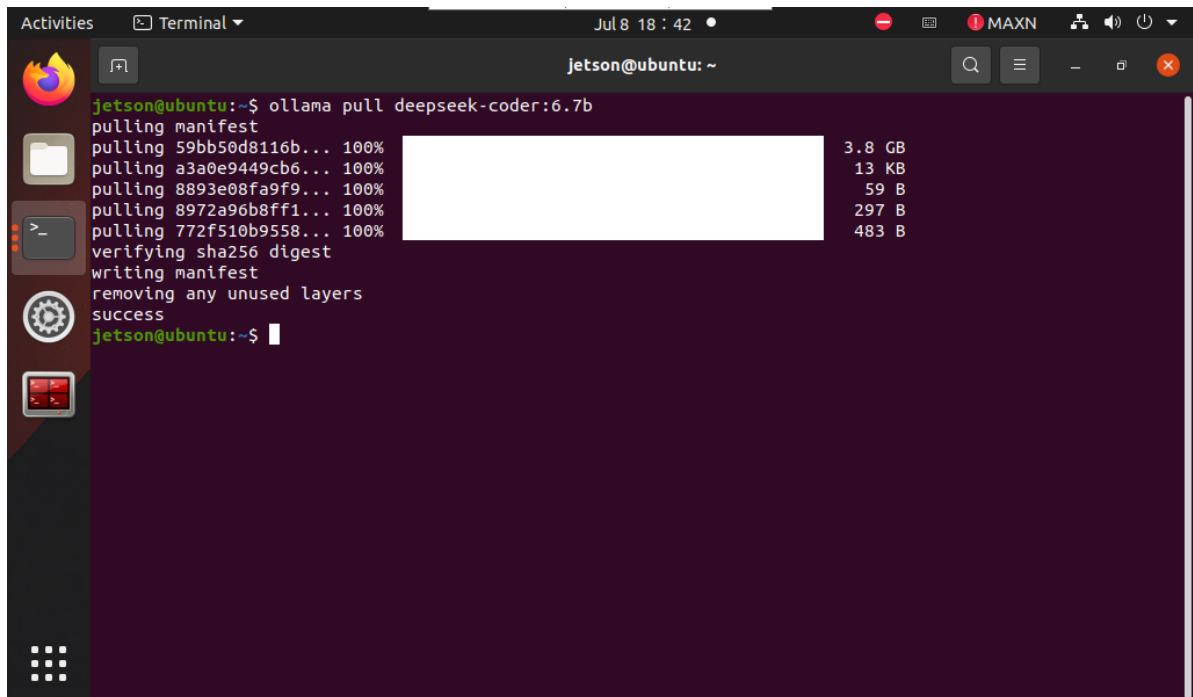
Pull DeepSeek Coder

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

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

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

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

A terminal window on a Jetson Ubuntu system showing the execution of the command 'ollama pull deepseek-coder:6.7b'. The output shows the model being pulled in layers: manifest, 59bb50d8116b... (3.8 GB), a3a0e9449cb6... (13 KB), 8893e08fa9f9... (59 B), 8972a96b8ff1... (297 B), and 772f510b9558... (483 B). The process includes verifying the sha256 digest and writing the manifest, resulting in success.

```
jetson@ubuntu:~$ ollama pull deepseek-coder:6.7b
pulling manifest
pulling 59bb50d8116b... 100% 3.8 GB
pulling a3a0e9449cb6... 100% 13 KB
pulling 8893e08fa9f9... 100% 59 B
pulling 8972a96b8ff1... 100% 297 B
pulling 772f510b9558... 100% 483 B
verifying sha256 digest
writing manifest
removing any unused layers
success
jetson@ubuntu:~$
```

Use DeepSeek Coder

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

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

Start a conversation

```
print HelloWorld in python
```

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

A terminal window titled 'jetson@ubuntu: ~' showing the process of pulling and running the 'deepseek-coder:6.7b' model. The 'ollama pull' command shows progress bars for pulling the manifest and several layers (59bb50d8116b, a3a0e9449cb6, 8893e08fa9f9, 8972a96b8ff1, 772f510b9558), all at 100%. It then verifies the sha256 digest, writes the manifest, and removes any unused layers, resulting in 'success'. The 'ollama run' command is then executed, followed by the user input 'print HelloWorld in python'. The model responds with a Python code snippet:

```
python
print("Hello World")
```

 and an explanation of the code. The prompt 'Send a message (/? for help)' is visible at the bottom.

```
jetson@ubuntu:~$ ollama pull deepseek-coder:6.7b
pulling manifest
pulling 59bb50d8116b... 100% [REDACTED] 3.8 GB
pulling a3a0e9449cb6... 100% [REDACTED] 13 KB
pulling 8893e08fa9f9... 100% [REDACTED] 59 B
pulling 8972a96b8ff1... 100% [REDACTED] 297 B
pulling 772f510b9558... 100% [REDACTED] 483 B
verifying sha256 digest
writing manifest
removing any unused layers
success
jetson@ubuntu:~$ ollama run deepseek-coder:6.7b
>>> print HelloWorld in python
Sure, here is a simple Python program that prints "Hello World":

```python
print("Hello World")
```

This code creates a print statement which outputs the string "Hello World" when it's run. In Python, 'print()' function is used to output data to the console. The string "Hello World" is passed as an argument to this function. When you run this program, it will display "Hello World" in your console.

>>> Send a message (/? for help)
```

End the conversation

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

A terminal window showing the continuation of the previous session. The user enters `/bye` at the prompt, which is highlighted with a red box. A second terminal window is overlaid on top, showing the model's response: `Use Ctrl + d or /bye to exit.`, which is also highlighted with a red box. The prompt 'Send a message (/? for help)' is visible in the background terminal.

```
jetson@ubuntu:~$ ollama pull deepseek-coder:6.7b
pulling manifest
pulling 59bb50d8116b... 100% [REDACTED] 3.8 GB
pulling a3a0e9449cb6... 100% [REDACTED] 13 KB
pulling 8893e08fa9f9... 100% [REDACTED] 59 B
pulling 8972a96b8ff1... 100% [REDACTED] 297 B
pulling 772f510b9558... 100% [REDACTED] 483 B
verifying sha256 digest
writing manifest
removing any unused layers
success
jetson@ubuntu:~$ ollama run deepseek-coder:6.7b
>>> print HelloWorld in python
Sure, here is a simple Python program that prints "Hello World":

```python
print("Hello World")
```

This code creates a print statement which outputs the string "Hello World" when it's run. In Python, 'print()' function is used to output data to the console. The string "Hello World" is passed as an argument to this function. When you run this program, it will display "Hello World" in your console.

>>> /bye
jetson@ubuntu:~$
```

```
jetson@ubuntu:~$ ollama run deepseek-coder:6.7b
>>> Use Ctrl + d or /bye to exit.
>>>
jetson@ubuntu:~$
```

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>

