Qwen2

> **Demonstration environment**

Development board: rdk x5 motherboard

sd card: 64G

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

Motherboard model	Ollama direct operation
rdk x5 8GB	√
rdk x5 4GB	√（需运行小参数版本）

Alibaba Qwen2 is an advanced open source large-scale language model developed by Alibaba, designed to provide powerful natural language processing capabilities.

1、Model scale

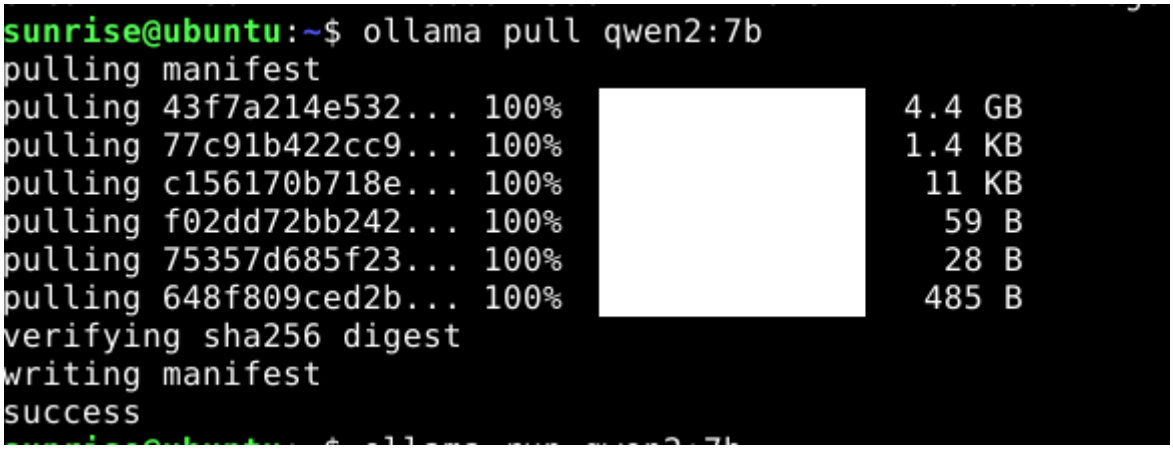
Model	Parameter
Qwen2	0.5B
Qwen2	1.5B
Qwen2	7B

2、Performance

3. Pull Qwen2

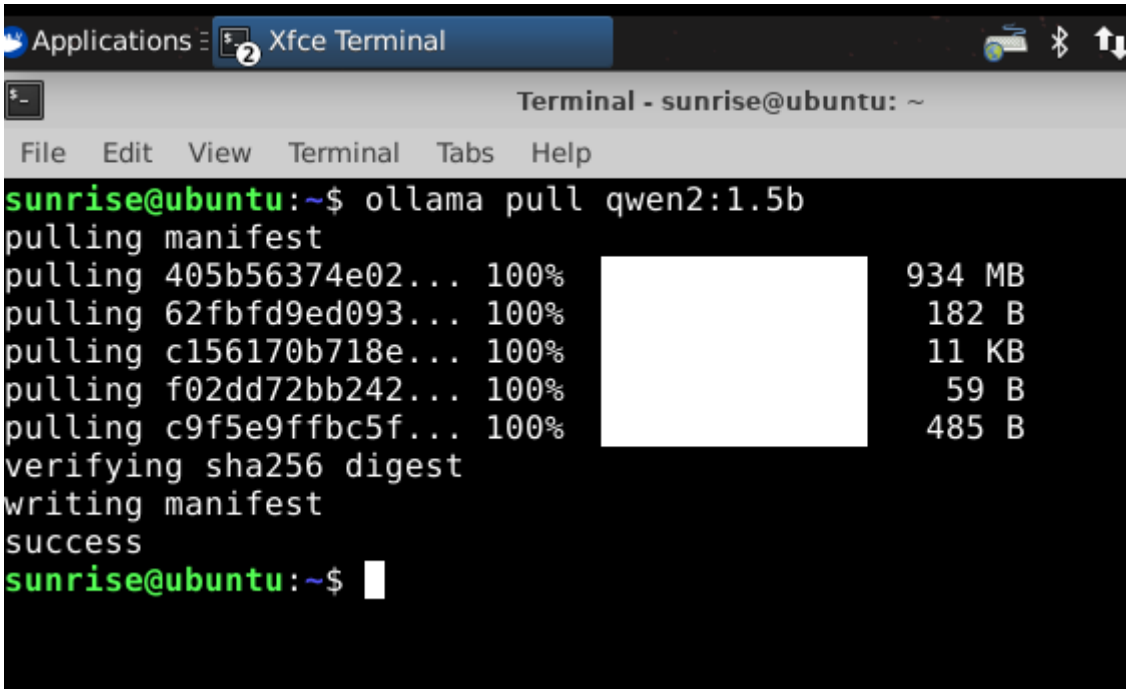
Using the pull command will automatically pull the model from the Ollama model library.

```
ollama pull qwen2:7b
```



Small parameter version model.

```
ollama pull qwen2:1.5b
```



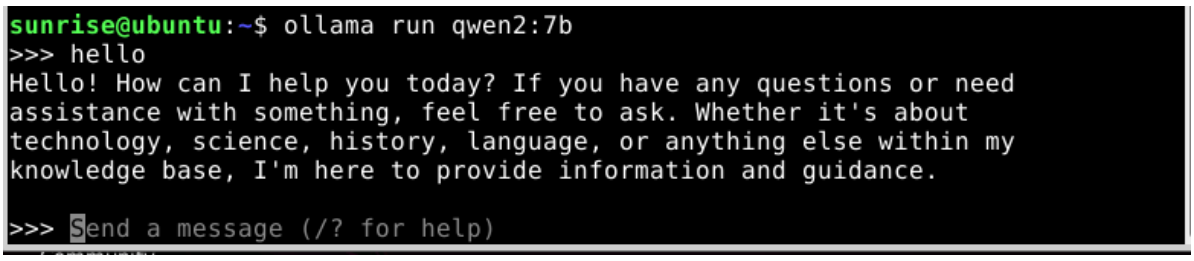
4. Use Qwen2

4.1. Run Qwen2

8gb motherboard runs the following parameters and smaller models:

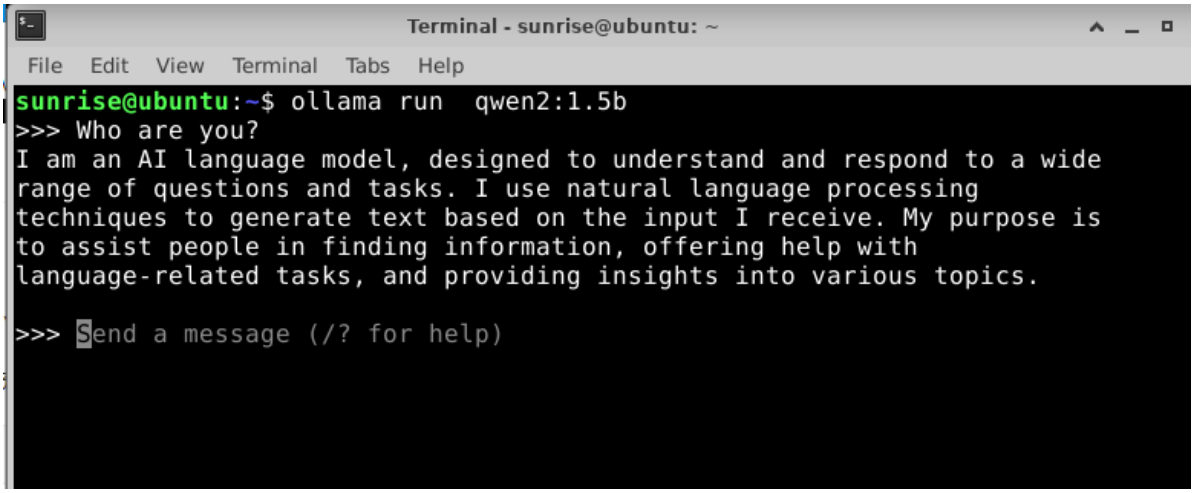
If the system does not have a running model, the system will automatically pull the Qwen2 7B model and run it:

```
ollama run qwen2:7b
```



A motherboard with 4 GB of RAM can run this model and smaller models.

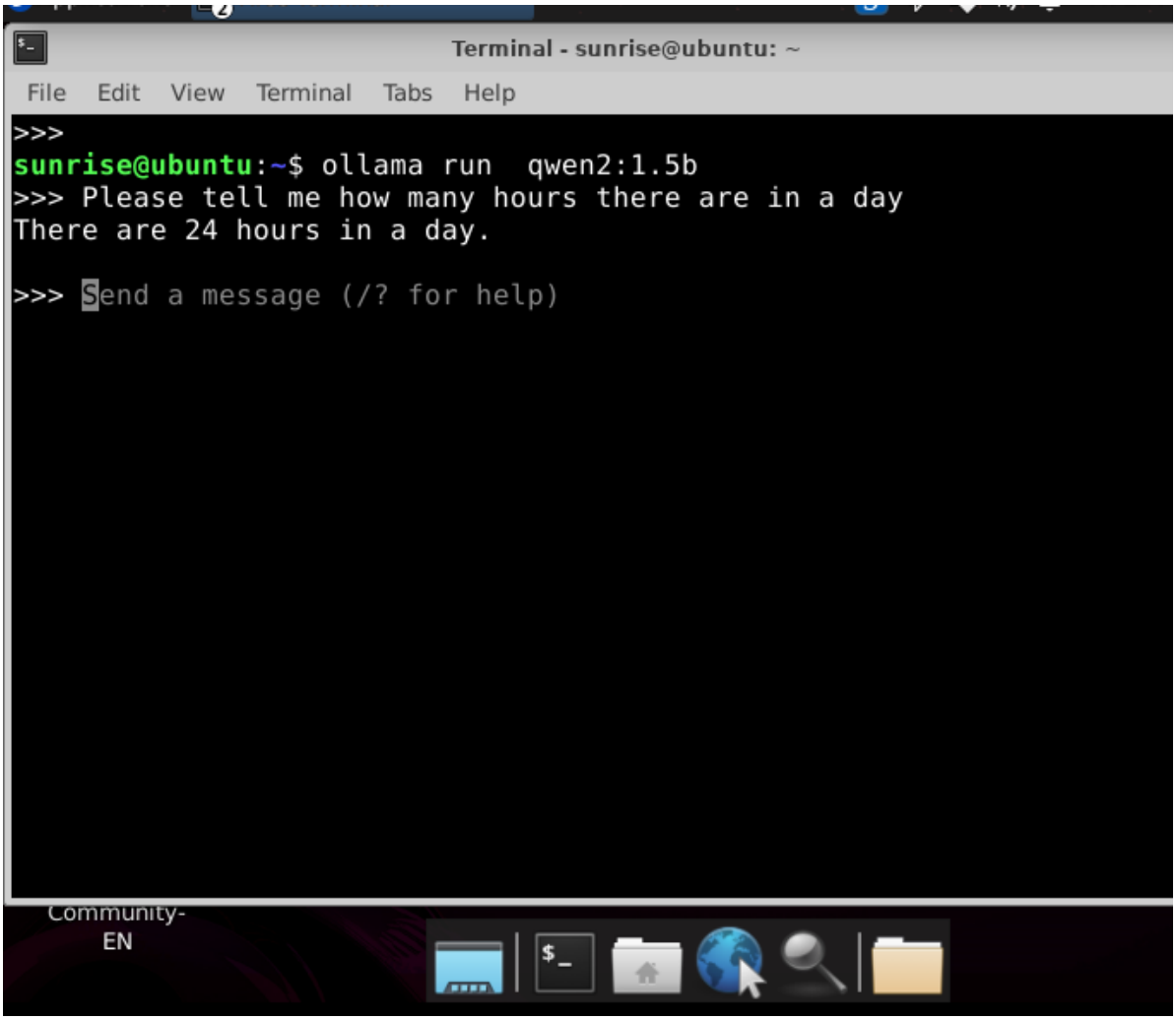
```
ollama run qwen2:1.5b
```



4.2. Have a conversation

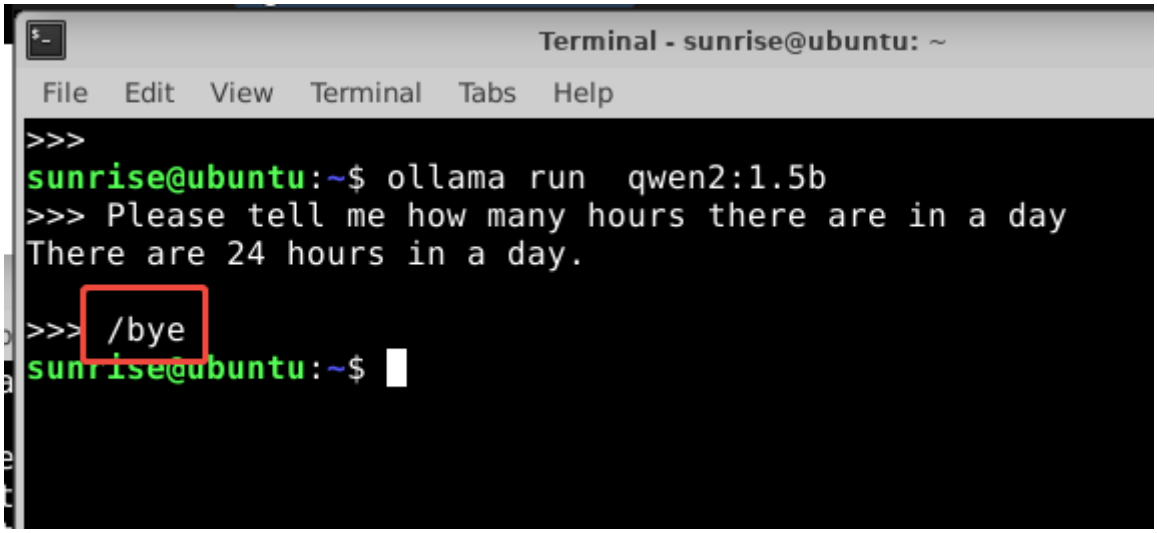
```
Please tell me how many hours there are in a day
```

The response time is related to the hardware configuration, please be patient.



4.3. End the conversation

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



References

> **Ollama**

Website：https://ollama.com/

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

> **Qwen2**

GitHub：https://github.com/QwenLM/Qwen2

Ollama model：https://ollama.com/library/qwen2