Model Usage Instructions

This tutorial mainly introduces some models that can be run in different motherboard image environments and different memory versions, as well as precautions.

Jetson Orin series can use the factory image directly.

Jetson Nano and Raspberry Pi 5 versions need to use the configured image provided in the network disk:

Jetson Nano: Jetson_Nano_AI_Pure.img

Raspberry Pi 5: Pi5_AI_Pure.zip

Jetson Orin series

Image description

Jetson Orin series offline AI large model development tutorial can use the factory image

> **System information**

User name: jetson

User password: yahboom

Hotspot name: ROSMASTER

Hotspot password: 12345678

> **System environment**

- Ollama
- Llama3
- Qwen2
- Phi-3
- Gemma
- WizardLM2
- Code Llama
- DeepSeek Coder
- Orca Mini
- StarCoder2

- TinyLlama
- LLaVA

LLaVA-Phi3

- Docker
- Open WebUI

Parameter Description

Jetson Orin NX 16GB, Jetson Orin NX 8GB, Jetson Orin Nano 8GB Can run models above 7B.

Jetson Orin Nano 4GB Runs small parameter models below 7B.

Although the above conclusions are not completely accurate, they can be used as a reference!

```
Ollam official website running model description: Running 7B parameter model
requires at least 8G available running memory, and running 13B parameter model
requires at least 16G available running memory.
```

Jetson Nano

Image Description

The offline AI large model development tutorial of Jetson Nano can use the image we configured: Jetson_Nano_AI_Pure.img

> **System Information**

Username: jetson

User password: yahboom

Hotspot name: Jetson_Hot

Hotspot password: 12345678

Motherboard IP address: 192.168.1.11

> **System Environment**

- Ollama
- TinyLlama
- Qwen2
- Phi-3
- Gemma
- Orca Mini
- Code Llama
- StarCoder2
- DeepSeek Coder
- LLaVA-Phi3
- Docker
- Open WebUI

Parameter Description

Jetson nano: Run 4B and below parameter models

Although the above conclusions are not completely accurate, they can be used as a reference!

Raspberry Pi 5

Image Description

The offline AI large model development tutorial of Raspberry Pi 5 can use the image we configured: Pi5_AI_Pure.img

> **System Information**

Username: pi

User Password: yahboom

Hotspot Name: Pi_Hot

Hotspot Password: 12345678

> **System Environment**

- Ollama
- Llama3
- Qwen2
- Phi-3
- Gemma
- WizardLM2
- Code Llama
- DeepSeek Coder
- LLaVA
- Docker
- Open WebUI

```
Due to the limitation of SD (TF) card, we did not download all the models of the tutorial to the Pi5_AI_Pure.img image system, and deleted the large parameter version models in some tutorials. Users can delete and download the models by themselves
```

Parameter Description

Raspberry Pi 5B (16G RAM): Run 14B and below parameter models

Raspberry Pi 5B (8G RAM): Run 8B and below parameter models

Raspberry Pi 5B (4G RAM): Run 3B and below parameter models

Raspberry Pi 5B (2G RAM): Run 0.5B and below parameter models

Although the above conclusions are not completely accurate, they can be used as a reference!

```
Ollam official website running model description: Run 7B parameter model at least 8G available running memory, run 13B parameter model model at least 16G available running memory, run 33B parameter model model at least 32G available running memory.
```

Model statistics

The following table is based on the common models counted by the Ollama official website model library. Users can find the required models on the Ollama official website.

```
Theoretically, the larger the model parameters, the more memory is required. You can choose a version with smaller parameters to experience the function when running the model.
```

Model	Parameters	Raspberry Pi 5B (2G RAM)	Raspberry Pi 5B (4G RAM)	Raspberry Pi 5B (8G RAM)	Raspberry Pi 5B (16G RAM)	Run Command
Llama 3	7B	Not supported	Not supported	Supported	Supported	ollama run llama3:8b
Llama 2 Chinese	7B	Not supported	Not supported	Supported	Supported	ollama run llama2-chinese:7b
Qwen2	0.5B	Supported	Supported	Supported	Supported	ollama run qwen2:0.5b
Qwen2	1.5B	Not supported	Supported	Supported	Supported	ollama run qwen2:1.5b
Qwen2	7B	No support	No support	Support	Supported	ollama run qwen2:7b
Phi 4	14B	Not support	Not support	Not support	Support	ollama run phi4
Phi-3	3.8B	No support	No support	Support	Supported	ollama run phi3:3.8b
Phi-2	2.7B	No support	Support	Support	Supported	ollama run phi:2.7b
Gemma	2B	No support	Support	Support	Supported	ollama run gemma:2b
Gemma	7B	No support	No support	Support	Supported	ollama run gemma:7b
LLaVA	7B	No support	No support	Support	Supported	ollama run llava:7b
WizardLM-2	7B	No support	No support	Support	Supported	ollama run wizardlm2:7b
DeepSeek Coder	1.3B	No support	Support	Support	Supported	ollama run deepseek-coder:1.3b
DeepSeek Coder	6.7B	Not supported	Not supported	Supported	Supported	ollama run deepseek-coder:6.7b
Yi	6B	Not supported	Not supported	Supported	Supported	ollama run yi:6b

References

> **Ollama**

Official website: https://ollama.com/

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