AI Large Model Image Usage Instructions

This tutorial mainly introduces the differences between the motherboard AI large model image and the car's factory image.

Important Notes

1. The default image in the TF card received with the car package is the car's factory image, only applicable to the AI large model development chapter. To run the examples in this chapter, please burn the corresponding image for your motherboard:

- RDK X5: Attachment\AI Large Model Image\RDK_X5\RDK_X5_AI_Pure.img
- Orin motherboards can directly run the examples in this chapter.
- Raspberry Pi 5: Attachment\AI Large Model Image\Raspberry Pi 5\Pi5_AI_Pure.img

2. Because Ollam requires a lot of memory to run large models, motherboards with limited memory can only use models with small parameters, or even cannot run them. Sometimes, when memory usage is at its limit, the performance will be poor. For a better experience of this feature, please refer to the online large model chapter. **
3. The commands in the AI large model section of the Raspberry Pi 5 are mostly run within Docker containers. You need to enter a Docker container before running any commands. This is not necessary for RDK X5 and Orin motherboards. Slowing down may occur when running within a Docker container; this is normal.

4. The ros2_docker.sh script on the Raspberry Pi 5 creates a container on its first run. If you don't manually delete this container, subsequent runs of the script will continuously enter it. If you encounter any issues or need to modify the Docker script, please delete the existing Docker container before rerunning the script. **

> **System Information**

- Raspberry Pi 5

Based on Raspberry Pi OS (64-bit): 2025-05-13

Username: pi

Password: yahboom

Hotspot Name: Pi_Hot

Hotspot Password: 12345678

Hotspot IP: 192.168.1.11

- RDK X5

Based on rdk_os_3.3.1-2025-08-08

Username: sunrise

Password: sunrise

- ORIN

Based on jetpack 6.2, ubuntu 22.04

Username: jetson

Password: yahboom

> **System Environment**

- Raspberry Pi 5, RDK X5, ORIN
- Ollama
- Docker
- Open WebUI
- Offline large model storage path: /usr/share/ollama/.ollama/models

Parameter Description

- Raspberry Pi 5

Raspberry Pi 5B (16GB RAM): Runs models with 14B or fewer parameters

Raspberry Pi 5B (8GB RAM): Runs models with 8B or fewer parameters

Raspberry Pi 5B (4GB RAM): Runs models with 3B or fewer parameters

Raspberry Pi 5B (2GB RAM): Runs models with 1B or fewer parameters

- RDK X5

RDK X5 (8GB RAM): Runs models with 8B or fewer parameters

RDK X5 (4GB RAM): Runs models with 3B or fewer parameters

While the above conclusions are not entirely accurate, they can be used as a reference!

Ollam's official website states that the following specifications apply to running models: 7B parameter models require at least 8GB of available RAM; 13B parameter models require at least 16GB of available RAM; and 33B parameter models require at least 32GB of available RAM.