AI large model system image instructions

- This tutorial mainly explains the differences between the factory images for the AI Large Model package and those for other packages.

Important Preface

The Jetson Nano AI Large Model Development Tutorial requires the pre-configured large model image: Motherboard/Appendix/Images/Factory Image_Online+Offline Large Model/AI_pure_Nano_4G_20250915.zip

1. If you purchased the AI Large Model package, the default image on the TF card you received is the AI Large Model image, which is only suitable for running the AI Large Model Development chapter. To run examples in other chapters, please burn the image in the "Factory Image_ROS+AI Vision" folder.

2. Because Ollama requires a large amount of memory to run large models, motherboards with limited memory can only use models with small parameters or may not run at all. Sometimes, performance may be poor when running at the maximum memory usage. For a better experience with this feature, please refer to the Online Large Model chapter.

The commands in the 3. Large AI Model section are mostly run within Docker. Before running these commands, you must first enter the Docker container. You may experience some lag when running within the Docker container, which is normal.

> **System Information**

Based on Jetson Nano Jetpack 4.6.3, Ubuntu 18.04

Username: jetson

User Password: yahboom

> **System Environment**

- Ollama
- Docker
- Open WebUI
- Offline Large Model Storage Path: /usr/share/ollama/.ollama/models

```
Due to the limitation of SD (TF) card, we did not download all the models in the tutorial to the AI_pure_Nano_4G_20250915.zip image system. We deleted the large parameter version models in some tutorials. Users can delete and download the models by themselves.
```

Parameter Description

Jetson Nano (4GB RAM): Runs models with 3B parameters or less

While the above conclusions are not completely accurate, they can serve as a reference!

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