AI large model system image instructions

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

Important Preface

This chapter requires the use of our pre-configured large model image, which can be obtained from the AI_pure_systemfile download section.

1. **The default image on the TF card received with the car package is the factory image for the car, only suitable for running the AI large model development chapter. To run the examples in this chapter, please burn the AI_pure_Nano_4G_20250915.img image file located in the Jetson nano folder.**
2. **Because Ollam requires a significant amount of memory to run large models, motherboards with limited memory can only use models with small parameter counts, or may not be able to run them at all. Sometimes, when memory usage is at its limit, the performance will be poor. For a better experience, please refer to the online large model chapter.**
3. **The commands in the AI large model chapter are mostly run within Docker containers. Before running commands, you need to enter a Docker container. Slowing down may occur when running within a 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.
```