RKLLM-Server-Flask. This facilitates quick replacement for users who have previously used OpenAI-API. In the subsequent sections of this chapter, we will separately introduce the one-click deployment script for the server, important settings for server deployment implementation, and the script design for client-side API access.

3.4.1.1 Server-side: One-click Deployment Script for RKLLM-Server-Flask

The one-click deployment script for RKLLM-Server-Flask is named build_rkllm_server_flask.sh and is located in the rkllm_server_demo directory. This script helps users quickly set up the RKLLM-Server-Flask server on a Linux development board. Before using this script, users should note the following:

1) Ensure that the development board is connected to the network via an Ethernet cable. Use the "ifconfig" command in the adb shell to determine the specific IP address of the development board. The RKLLM-Server-Flask will then set up the server with this IP address within the local area network and accept client access.

2) Users should have successfully converted the RKLLM model beforehand. Before executing the one-click deployment script, ensure that the RKLLM model has been pushed to the Linux board.

Users can directly invoke the build_rkllm_server_flask.sh script from their PC (not on a development board) to quickly deploy the RKLLM-Server-Flask server on a Linux development board.

The specific usage of the one-click deployment script build_rkllm_server_flask.sh is as follows:

```
./build_rkllm_server_flask.sh
   --workshop [RKLLM-Server Working Path]
   --model_path [Absolute Path of Converted RKLLM Model on Board] --
platform [Target Platform: rk3588/rk3576]
   [--lora_model_path [Lora Model Path]]
   [--prompt_cache_path [Prompt Cache File Path]]
```

The `workshop` parameter specifies the subsequent working directory of RKLLM-Server-Flask on the board. The `model_path` parameter indicates the absolute path of the RKLLM model on the board, which was converted using RKLLM-Toolkit, and RKLLM-Server-Flask will read the model from this path during operation. The `platform` parameter specifies the current platform type, either rk3588 or rk3576. The `lora_model_path` and `prompt_cache_path` parameters are optional and can be used to