

setup of RKLLM-Server-Gradio on a Linux development board. Similar to the setup process for RKLLM-Server-Flask, users should pay attention to the following points before using the one-click deployment script:

1) Ensure that the development board is connected to the network via an Ethernet cable. Use the ifconfig command in the adb shell to query the specific IP address of the development board. RKLLM-Server-Gradio will be set up as a server within the local network using this IP address to accept client access.

2) Users need to complete the smooth conversion of the RKLLM model and have pushed the RKLLM model to the Linux board before executing the one-click deployment script.

The usage of the one-click deployment script build\_rkllm\_server\_gradio.sh is similar to that of build\_rkllm\_server\_flask.sh:

```
./build_rkllm_server_gradio.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]]
```

Similarly, the workshop parameter specifies the working directory for RKLLM-Server-Gradio on the device; the model\_path parameter indicates the absolute path of the RKLLM model on the device, which was converted using RKLLM-Toolkit, and RKLLM-Server-Gradio will read the model from this path during operation; the platform parameter specifies the platform type being used, such as rk3588 or rk3576; lora\_model\_path and prompt\_cache\_path are optional parameters, allowing the user to specify the file paths for loading a Lora model or utilizing the Prompt feature if needed.

Users can directly call the build\_rkllm\_server\_gradio.sh script on the PC side (not on the development board) using the following command to quickly deploy the RKLLM-Server-Gradio example:

```
./build_rkllm_server_gradio.sh
--workshop /user/data
--model_path /user/data/model.rkllm
--platform rk3588
```

After executing the above command, the one-click deployment script will perform the following steps: