

|         | 3.2.3 on input structure definition.   |
|---------|--|
|         | RKLLMInferParam rkllm_infer_params*: Parameter passing during the model              |
|         | inference process. For details, see section 3.2.3 on input structure definition.     |
|         | void* userdata: the user-defined function pointer, typically set to NULL by default. |
| Returns | 0 indicates that the model inference runs normally;                                  |
|         | -1 indicates a failure in calling the model inference;                               |

The example code is as follows:

```
#define PROMPT TEXT PREFIX "<|im start|>system You are a helpful
assistant. <|im end|> <|im start|>user"
#define PROMPT TEXT POSTFIX "<|im end|><|im start|>assistant"
// Predefined text values for the prompt before and after
string input str = "把这句话翻译成英文: RK3588 是新一代高端处理器, 具有高算力、
低功耗、超强多媒体、丰富数据接口等特点";
input str = PROMPT TEXT PREFIX + input str + PROMPT TEXT POSTFIX;
// Define the input prompt and complete the concatenation
RKLLMInferParam rkllm_infer_params;
memset(&rkllm infer params, 0, sizeof(RKLLMInferParam));
rkllm infer params.mode = RKLLM INFER GENERATE;
// 1. Initialize and set LoRA parameters (if needed)
RKLLMLoraParam lora params;
lora params.lora adapter name = "test";
// 2. Initialize and Set Prompt Cache Parameters(if needed)
RKLLMPromptCacheParam prompt cache params;
prompt cache params.save prompt cache = true;
prompt cache params.prompt cache path = "./prompt cache.bin";
rkllm infer params.mode = RKLLM INFER GENERATE;
// rkllm infer params.lora params = &lora params;
// rkllm infer params.prompt_cache_params = &prompt_cache_params;
rkllm infer params.lora params = NULL;
rkllm infer_params.prompt_cache_params = NULL;
RKLLMInput rkllm input;
rkllm_input.input_type = RKLLM_INPUT_PROMPT;
rkllm_input.prompt_input = (char *)text.c str();
rkllm run(llmHandle, &rkllm input, &rkllm infer params, NULL);
```

## 3.2.6 Interrupt Model Inference

During model inference, users can call the rkllm\_abort() function to interrupt the inference process.

The specific function definition is as follows:

Table 3-22 Interface Specification for the rkllm abort Function