

Definition	RKLLMResult
Introduction	Used to return the current inference-generated result.
Struct Fields	<p>const char* text: indicates the text content generated by the current inference;</p> <p>int32_t token_id: indicates the token_id generated by the current inference;</p> <p>RKLLMResultLogits logits: Represents the logits information generated during the current inference, only returned when RKLLMInferMode is set to RKLLM_INFER_GET_LOGITS;</p>

3) RKLLMResultLogits is a return value structure, and its specific definition is as follows:

Table 3-9 Explanation of RKLLMResultLogitsStructure

Definition	RKLLMResultLogits
Introduction	Used to return the current inference-generated logits.
Struct Fields	<p>const float* logits: indicates the logits generated by the current inference;</p> <p>int vocab_size: indicates the token_id generated by the current inference;</p> <p>int num_tokens: Indicates the number of tokens returned; users can calculate the size of the logits based on vocab_size and num_tokens;</p>

During the design process of the callback function, users can set different post-processing behaviors based on the values in RKLLMResult.

3.2.2 Define RKLLMParam

The RKLLMParam structure is used to describe and define the detailed information of RKLLM.

The specific definition is as follows:

Table 3-10 Explanation of RKLLMParam Structure

Definition	RKLLMParam
Introduction	Used to define the detailed parameters of the RKLLM model.