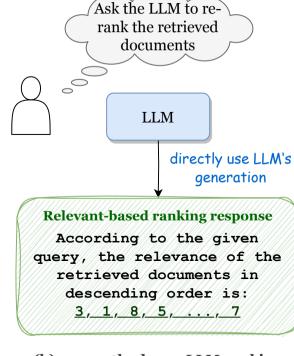


(a) previous work rely on LLM's likelihood to directly distill training the retriever



(b) our method uses LLM ranking response to intermediately distill training the retriever