1.什么是RAG

RAG--Retrieval-Augmented Generation（检索增强生成），是一种混合模型架构,是一种结合了信息检索和序列到序列的文本生成技术。

2.RAG是哪一年提出的概念，在什么背景下诞生了RAG

RAG的概念是在2020年由Facebook AI Research团队提出。

为了克服言生成模型在提供最新信息或领域专业知识方面存在局限性问题，研究人员探索了将信息检索与语言生成相结合的方法，从而提出了RAG。

3.什么是KG和LLM

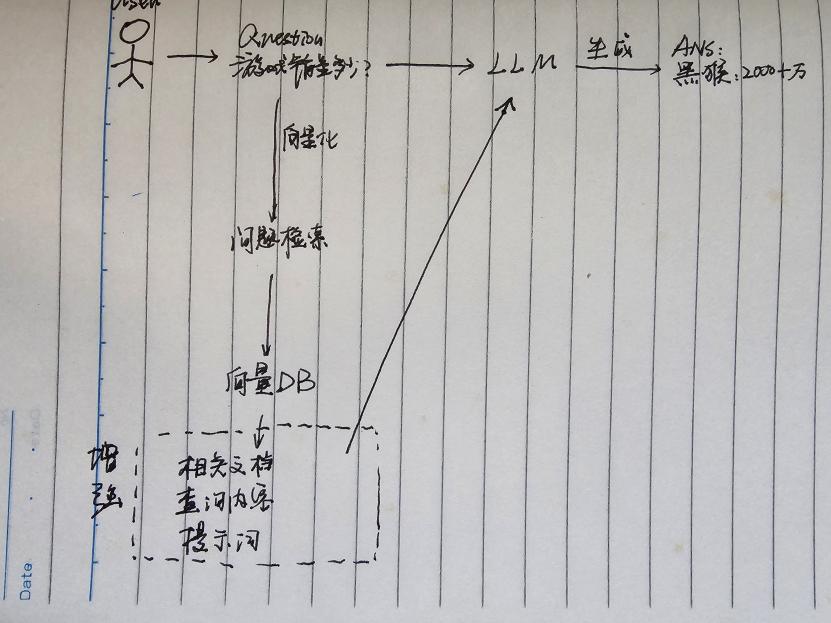
KG：知识图谱，一种图形数据结构表示现实世界中的概念及其之间的关联。

LLM ：大语言模型，指那些具有大量参数的深度学习模型。作用于语言的回应与生成

4. RAG的工作原理是什么

RAG的工作原理是通过结合信息检索和生成模型，先从大型文档集合中找到与问题最相关的文档片段，再利用这些片段生成精准的回答。

1. 自己在草稿纸上画一张RAG工作流程图



6. 现在的RAG已经用了什么技术，解决了什么场景的问题

技术：TensorFlow, PyTorch

增强聊天机器人对复杂话题的理解和回应能力。

7. 现在的RAG面临什么挑战，在哪方面做的有欠缺

现在的RAG面临的主要挑战包括处理多模态数据的复杂性和确保生成内容的一致性与逻辑性。此外，在低资源语言支持和高效检索大规模数据集方面仍存在不足。

1. RAG技术演进历程，现在最新的技术

截至2024年底，RAG技术的最新进展包括GraphRAG、多模态处理、混合检索增强、自适应提示工程以及强化学习的应用，显著提升了模型在复杂场景下的准确性和灵活性。

1. RAG在政策问答上的研究论文列举出来，已经实现的应用的GitHub链接

RAGflow:https://github.com/infiniflow/ragflow

关于RAG的文答以及模型的问答：https://direct.mit.edu/tacl/article/doi/10.1162/tacl\_a\_00530/114590/Improving-the-Domain-Adaptation-of-Retrieval