



GLM4 is all you need

使用GLM4从0到1构建并优化你的rag程序

2024年2月 智谱AI技术团队

北京智谱华章科技有限公司

RAG（检索增强式生成）

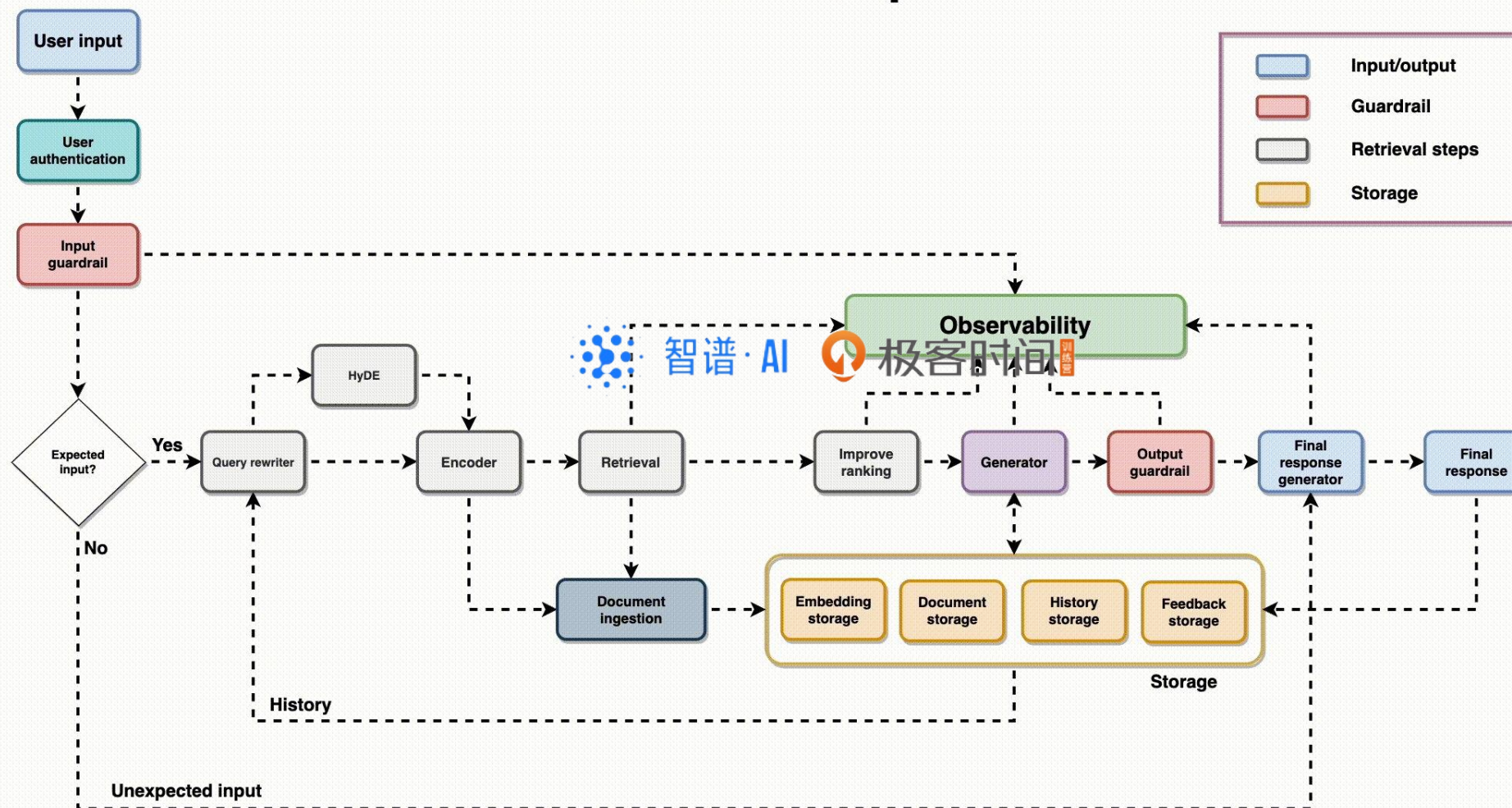
什么是RAG？

RAG，即检索增强式生成（Retrieval Augmented Generation），是一种结合了检索和生成两种方法的自然语言处理技术。它利用检索技术从大量数据中快速找到相关信息，然后使用生成技术对这些信息进行整合和创造，生成高质量的文本。

RAG的作用

1. **提高生成文本的质量和准确性：**通过检索技术，RAG可以确保生成的内容基于可靠的信息源，从而提高文本的准确性和质量。
2. **增强生成文本的多样性：**RAG可以从多个信息源中检索信息，使得生成文本的内容更加丰富和多样化。
3. **提高生成速度：**检索技术可以快速定位到相关信息，从而加快生成过程。
4. **实现长文本生成：**RAG可以处理长文本生成任务，因为它可以从多个信息源中检索和整合信息，生成连贯且内容丰富的长文本。

Architecture For Enterprise RAG



GLM4 is all you need - 使用GLM4从0到1搭建并优化你的rag程序

1. 使用GLM4清洗数据
2. 使用GLM4来结合rag生成回答
3. 使用GLM4来评估你生成的回答
4. 使用GLM4来标注你的检索数据
5. 使用GLM4标注数据训练你的Adapter，效果拔群！