检索增强生成

知识检索 · 基于证据生成

检索增强生成先从外部资料中检索相关内容，再把检索结果交给生成模型形成回答。知识库可以独立更新，通常不需要为每次资料更新重新训练模型。

01 工作方式

01 整理资料，保留文档来源、版本与访问权限。

02 按问题检索相关段落，必要时重新排序。

03 依据检索证据生成答案，同时给出可回溯引用。

02 场景应用

01 工艺文件、设备手册和内部制度问答。

02 基于授权院内资料的知识检索与说明。

03 边界与常见误区

01 检索错误、资料过期和生成错误仍然可能发生。

02 检索增强生成不能消除幻觉，也不自动证明资料内容正确。

03 资料权限应在检索与回答环节贯彻，不能只隐藏界面入口。

04 来源与阅读说明

依据下列一手资料原创归纳；场景与工程建议为编辑整理，不是来源机构对本站服务的背书。

国际商业机器公司官方资料 · 什么是检索增强生成

https://www.ibm.com/think/topics/retrieval-augmented-generation