

周凯

Tel: 15995397689

Email: kaizhou0305@gmail.com

WeChat: zk471104594

Blog: zhoukai.space



教育经历

吉林大学

人工智能专业硕士

吉林/长春

2025.9 - 2028.7

GPA: 3.4/4.0 |

南京信息工程大学

数据科学与大数据技术本科

江苏/南京

2021.9 - 2025.7

GPA: 3.4/5.0 | 校优秀学生干部, 国家励志奖学金, 校奖学金等

项目经历

多模态专利文档智能问答系统

项目简介: 基于阿里云天池竞赛数据集构建的多模态专利文档问答系统，通过检索增强生成 & 大模型微调实现专利文档的智能理解与精准回答。

技术栈: Transformers、PyMuPDF、TRL、PEFT、LoRA、Qwen、PyTorch、Datasets、RLHF、vLLM

1) 多模态数据预处理: 使用 PyMuPDF 将专利 PDF 转换为高分辨率图像（600 DPI），基于 Alibaba-NLP/gme-Qwen2-VL-7B-Instruct 模型提取向量表示，构建问题、图像的向量数据库。

2) 智能检索系统: 问题 & 图像的跨模态语义检索，基于余弦相似度精准定位相关文档页面（Top-K=2），根据问题类型动态调整图像输入组合。

3) 模型微调: 构建差异化训练数据集（专利文档问答数据+少量问题分类辅助数据），缓解模型过拟合风险；基于 Qwen3-VL-32B 多模态大模型，采用 LoRA 完成参数高效微调；结合梯度累积、学习率调度等策略优化训练流程，显著提升模型在专利问答任务中的响应准确率与鲁棒性。

4) 差异化推理策略: 搭建问题分类模块，精准判别纯图像位置关系 & 推理混合两种问题类型；定制专属提示词，通过差异化输入引导策略提升初答准确性。

5) 答案优化机制: 设计“初答+风格迁移”两阶段优化流程，基于 Few-Shot 范式构建风格示例库（检索相似问题答案作为风格参考），通过后处理模块提取核心表达风格，提升答案的质量与一致性。

6) 性能提升: 通过微调 & 检索优化 & 后处理，最终 BERTScore 的 F1 从 0.6674 提升至 0.8452，EM 从 6.7% 提升至 24.52%，ROUGE-1 的 F1 从 0.2139 提升至 0.5495。

智能文献检索调研 Agent 系统

项目简介: 面向科研调研场景，构建具备任务规划、工具调用、多轮推理与结构化输出能力的多 Agent 文献调研系统。

技术栈: Python、LangChain、Agent Planning、Tool Calling、Vector Retrieval、Reranker

1) Agent 架构设计: 基于 LangChain 构建 Research Agent，负责调研目标解析、子任务拆分、执行调度与结果整合，形成 Plan → Execute → Review 的闭环推理流程。

2) 调研任务规划与自我修正: 根据用户主题生成结构化调研计划（关键词扩展、检索来源、筛选标准、输出结构），在检索结果不足或偏离目标时自动调整关键词与检索策略。

3) 相关性评估与筛选: 采用向量相似度完成候选文献初筛，引入 Reranker 对候选结果精排，压缩无关文献集合；基于排序结果判断是否进入解析阶段或触发补充检索。

4) 文献解析与理解: 使用 PyMuPDF 对 PDF 文档进行结构化切分，调用文献理解 Chain，抽取研究问题、方法、实验设置与核心结论等关键信息。

技能/证书/竞赛

证书: CET4、CET6

语言: 精通 Python 开发

技术: 熟悉 Transformer、BERT、QwenVL 等主流大语言模型原理与架构；熟练运用 LoRA 完成大模型参数高效微调；熟悉 DPO、PPO、GRPO 等主流 RL 对齐算法

应用: 熟悉 RAG 检索增强技术及主流优化方法；掌握 Agent 开发、工具调用全流程