

胡树海

15060396956 • 1955034593@qq.com • 男 • 22

教育经历

2022.9-2026.6

莆田学院

数据科学与大数据技术 | 本科

相关证书：英语四级证书

相关技能

- 编程语言：Python、Java
- 后端框架：FastAPI、Flask
- AI / Agent：LangChain、LangGraph、Function Calling、RAG、混合检索、多 Agent 编排、Prompt Engineering
- 深度学习：PyTorch、Transformer、BERT
- 爬虫：requests、BeautifulSoup4、Playwright、Scrapy 框架
- 工程化：pytest、Docker Compose、Git
- 智能体：claude code、codex

项目经历

AI 知识库问答系统 (DocQ)

2025.9-2026.3

项目介绍：基于 LangGraph 编排 6 节点 Agent 管道，实现路由判断→查询改写→混合检索→重排序→生成回答→幻觉检查全流程自动化。

- 技术栈**：Python 3.12、LangChain、LangGraph、FastAPI、ChromaDB、Streamlit、DeepSeek API。
- 设计向量 + 关键词混合检索架构——ChromaDB HNSW 索引负责语义召回，BM25 倒排索引负责精确匹配（函数名、配置项），RRF 算法融合双路排序，LLM 二次重排序精选 Top-2 文档。
- 文档摄入管道：PDF/Markdown 解析→按 Markdown 标题层级智能分块（代码块和表格不拆散）→DeepSeek Embedding 向量化→ChromaDB 持久化存储。
- 幻觉检查自校正——生成回答后逐条对照原文验证事实，发现编造内容自动触发重新生成，最多重试 2 次，保证回答忠实于文档。
- 设计查询改写节点——LLM 将用户口语化问题（“那个东西怎么配”）补全指代、对齐中英文术语，改写为精确检索关键词。

日志异常检测系统 (LogBERT)

2024.3-2024.9

项目介绍：基于 BERT 预训练模型构建日志异常检测系统，通过自监督学习捕捉正常日志模式，识别偏离正常分布的异常序列，提升复杂系统中故障排查效率。

- 技术栈**：Python、PyTorch、Flask、Transformer、Drain、Pandas、NumPy。
- 手写 Transformer 核心组件——多头注意力机制、前馈网络层、嵌入层，纯 PyTorch 实现，不依赖 HuggingFace。
- 掩码语言模型预训练——随机 Mask 日志序列中 15% token，训练模型从上下文推断被掩盖的日志键，预测概率低于阈值即标记为内容异常
- 超球面序列建模——将正常日志序列的向量在空间中的分布拟合为超球体，新序列的向量偏离球心距离超过半径时判定为模式异常
- Drain 固定深度解析树将非结构化日志聚类为模板序列，自建词表无需人工标注
- Flask 部署 /predict API——加载预训练模型权重，接收 logKey 序列实时推理，单次预测毫秒级返回。