

李赢洲

AI 应用开发实习 | Python / FastAPI / RAG / Agent / LangGraph

电话: 13054111853 | 邮箱: 13054111853@163.com | GitHub: github.com/20082123

个人网页: my-flow.top



电子信息硕士在读, 面向 AI 应用开发与工程化落地, 具备 Python/FastAPI、RAG、LangGraph Agent、文档解析、异步任务和模型调用优化实践。实习中参与工业文档智能问答 PoC; 个人项目重点验证多智能体意图路由、工具调用边界、Memory、Human-in-the-loop 与 RAG 评测闭环。

教育背景

贵州大学 2024.09-至今
电子信息 (硕士); 华为杯中国研究生数学建模竞赛国家二等奖; 校三等奖学金; 研究方向: 随机配置网络、强化学习
防灾科技学院 2020.09-2024.06
信息管理与信息系统 (本科); 河北省数据分析竞赛省三等奖; 两次校二等奖学金

专业技能

Python 后端开发: 熟悉 Python、FastAPI、asyncio、Pydantic、RESTful API、异常兜底和 Worker 异步处理, 能够将模型调用封装为可调试的后端接口与任务链路

RAG 与文档处理: 熟悉 RAG、文档解析清洗、语义分块、Prompt 约束、结构化输出、pdfplumber、PaddleOCR、Milvus、BM25/向量混合检索、Query Rewrite 和 ReRanker 思路

Agent 应用开发: 熟悉 LangGraph / LangChain、Conditional Edges、Tool Calling、意图识别、会话 Memory、任务路由和 Human-in-the-loop 等 Agent 核心机制与原型实现

工程化与评测: 熟悉 Docker Compose、Git、Redis、SQLite、RabbitMQ、日志分析、测试样例构造和 Bad Case 归因; 了解 Dify 工作流与 LoRA/P-Tuning 数据准备流程

实习项目经历

工业文档智能问答 PoC 上海波士内智能科技有限公司 | AI 应用工程实习 (3 个月)
面向工业 PDF 表格多、版式复杂、人工查询效率低的问题, 参与工业文档智能问答 PoC 的文档解析、异步任务、模型调用优化和图谱入库模块开发。

- 在导师指导下参与“逐页检测 + 动态路由”解析方案工程实现, 独立完成基于 pdfplumber 字符数阈值的低密度页面判定模块, 并负责 PaddleOCR 切换逻辑联调与测试验证; 配合整体方案在 7 页混合 PDF 样例上验证解析耗时由 21.3 分钟降至约 13 秒。
- 独立负责异步任务模块, 将长文档处理从同步请求改造为 FastAPI + SQLite (PoC 阶段轻量方案) + Worker 异步链路, 实现任务状态持久化、上传/状态查询/失败记录 REST 接口和失败重试机制, 降低长任务阻塞请求链路的风险。
- 针对 LightRAG 每次请求重复加载模型导致推理延迟高的问题, 实现模型客户端常驻缓存, 减少重复初始化与显存分配; 在单 Chunk 测试中验证推理延迟由 144s 降至 1.2s, VRAM 峰值占用降低 96%。
- 独立实现 Neo4j 批量写入脚本, 采用 UNWIND 写入节点与边, 完成 RAG 检索结果到图关系数据的转换与校验, 并处理特殊字符转义、重复关系合并等工程细节。

个人实践与项目

基于 LangGraph 的无人机集群多智能体指挥原型 个人项目 | LangGraph / Intent Routing / Tool Calling / HITL
面向无人机任务指挥中自然语言意图混杂、工具边界不清和敏感控制风险高的问题, 基于 LangGraph 搭建多智能体编排原型, 验证从意图识别、任务路由、工具调用到人工确认的 Agent 应用链路。

- 设计“总指挥 Agent + 航线规划 / 状态诊断 / 载荷任务 / 运维问答专员 Agent”的分层结构, 使用 LLM 意图识别与 Conditional Edges 将指令路由至对应节点; 针对复合指令引入任务拆解与执行顺序约束, 减少单体 Prompt 同时处理多类任务的混乱。
- 封装 get_drone_status、plan_route、check_battery、return_home、land 等模拟工具, 基于 Pydantic 校验工具入参与业务约束, 区分只读查询、规划建议和敏感控制动作, 便于后续替换为真实无人机 API。
- 维护 Thread 级会话状态, 记录当前无人机、任务目标、关键槽位和工具执行结果; 对返航、降落等高风险动作设置 Human-in-the-loop 中断确认, 并整理测试指令与 Bad Case 用于评估路由和工具调用稳定性。

基于 RAG 的无人机智能运维知识库原型 个人项目 | 文档解析 / 混合检索 / Query Rewrite / RAG 评估
面向无人机维修手册中故障代码、跨页表格和图文混排难检索的问题, 搭建文档解析、语义分块、混合检索和查询改写链路, 验证维修问答场景下的 RAG 检索增强效果。

- 使用 Unstructured / OCR 解析维修手册、故障码表和保养记录, 按章节层级、故障对象和语义边界进行分块, 保留文档来源、页码、部件型号等 metadata, 支撑答案溯源。
- 构建向量检索 + 关键词召回的混合检索链路, 针对故障代码、部件型号等精确词设计召回策略, 并引入 Query Rewrite 将口语化故障描述改写为更适合检索的问题。
- 设计 RAG 回答约束, 要求引用来源片段并输出故障原因、排查步骤、注意事项等结构化结果; 整理测试问答集和 Bad Case 表, 按未命中、错召回、答案不完整等类型迭代分块、检索和 Prompt。