

性别：男

工作年限：2 年

学历：统招本科

入职时间：一周内到岗

联系电话：18697763206

邮箱：18697763206@163.com

工作经历

七彩阳光（北京）旅行社有限公司 | AI-Agent 开发

2025.07 - 2026.07

易数通数据分析平台

项目职务：AI-Agent 开发 项目周期：2026.01 - 2026.07

技术栈：Python、FastAPI、LangChain、Milvus、MySQL、SSE、Redis、SQL AST 校验

项目简介：面向公司财务、采购、经营分析和管理层等角色，构建基于自然语言的经营分析系统，支持通过自然语言查询订单、客户、产品、收入和利润等数据，覆盖问题理解、指标识别、Schema Linking、SQL 生成、执行和结果分析的完整链路，辅助区域资源投放和经营决策。

项目职责：

- Text2SQL 链路开发**：针对业务人员通过自然语言查询经营数据的需求，负责搭建“输入校验 -> 指标 RAG -> Schema Linking -> Prompt 构建 -> LLM 生成 -> SQL AST 校验 -> 数据库执行 -> 结果解析”链路，支持动态 Schema 和结构化结果输出。
- 指标知识库与指标 RAG**：为统一经营分析口径，维护指标名称、别名、定义、公式、数据源和依赖关系；通过指标 RAG 语义检索，自动展开派生指标依赖，为 SQL 生成提供指标依据。
- Schema Linking**：针对 200+ 张表及大量字段带来的 Schema 上下文开销，实现表召回、字段匹配和 BFS Join 路径推理，将全量 Schema 注入调整为动态召回，Prompt Schema Token 降低约 55%。
- Plan-and-Execute 多步分析**：针对利润分析、费用异常、区域对比等复杂问题，将查询拆解为多个 SQL 子任务，通过依赖执行、中间结果管理和失败重试完成多步分析，复杂问题可执行率达到 91%。
- 评估体系与 Bad Case 分析**：为定位查询链路的薄弱环节，设计简单、中等、复杂三层测试集，基于 Execution Accuracy 和 Exact Match 评估，定位多表 Join、时间过滤和指标口径等问题。
- 服务化与运行保障**：为支持业务系统调用，基于 FastAPI 提供 RESTful API 和 SSE 流式输出，补充 SQL AST 校验、权限控制、异常分类和慢查询分析，保障查询链路的运行。
- 业务应用与效果**：以自助查询率 72% 为目标，支持业务人员自助完成经营数据查询和分析，数据团队重复取数需求下降约 80%，缩短经营分析问题的响应周期。

空铁无忧智能客服

项目职务：AI-Agent 开发 项目周期：2025.07 - 2025.12

技术栈：Spring Boot、Python、Redis、LangChain、Milvus、MySQL、RAG、Function Calling、Rule Engine

项目简介：面向头部 OTA 平台的机票及辅营产品售后场景，处理退票、改签、选座、餐食、行李等咨询。基于分层意图识别、RAG 政策检索、规则引擎和结构化结果生成，自动判断订单能否退改并计算费用等，减少人工查询和计算成本，帮助平台降低售后人工成本并提升处理效率。

项目职责：

- 意图识别与置信度路由**：设计“规则引擎 + 小模型 + 大模型兜底”的 Intent Pipeline，识别退票、改签、辅营等意图，并按置信度自动处理、追问澄清或转人工。
- Query 理解与改写**：结合多轮上下文、订单槽位通过 Query Rewrite 完成指代消解、省略补全与术语标准化，并分别用于意图路由和 RAG 检索，提升多轮召回质量。
- RAG 知识库构建**：完成航司政策、退改规则、辅营说明等文档的解析、Chunk 切片、Embedding 向量化和索引构建，并使用 Milvus 存储向量索引形成统一、可检索、可引用的政策知识库。
- 混合检索与重排优化**：采用 Hybrid Search (BM25 + 向量) 进行双路召回，通过 RRF 融合、MMR 去重、航司 / 产品 / 票种元数据过滤及 Rerank 重排，降低政策串用，召回准确率提升约 35%。
- 多轮会话与结构化输出**：采用 15 轮滑动窗口模式，通过 Redis 存储历史摘要、关键槽位和会话状态，避免重复描述，实现 Token 消耗降低约 44%，关键事实召回率 95%。
- 防幻觉与多源政策校验**：在 System Prompt 层实施 RAG 幻觉抑制，结合 temperature=0 与引用溯源，强制模型仅依据选定的航司政策或本地文档回答并输出 source；本地命中不足或时效性不足时联动官网校验，幻觉回答率由 30% 降至 4% 以下。
- 退改规则与费用计算**：基于规则引擎按“航司 × 产品 × 票种 × 订单状态”匹配政策，校验退改资格并计算手续费和退款金额，避免大模型直接参与业务决策。
- 业务效果与服务化**：将人工查询、计算流程优化为“系统自动判断 + 人工确认”，人工转接率下降 72%。

教育背景

2023.09 - 2025.06 郑州科技学院 软件工程（本科）

GPA 3.54 / 4 (专业前 10%)

主修课程：数据结构、操作系统、计算机网络、软件工程、数据库、Java Web、软件测试、Linux、计算机组成原理、人机交互等。

专业技能

AI 应用开发

- **LLM 应用与 Agent 开发**：熟悉 LangChain、LangGraph、Spring AI 等大模型开发框架；熟悉 Agent 构建、ReAct 机制、Function Calling、工具调用、记忆机制与多轮会话设计，能够完成知识增强应用与复杂业务逻辑落地。
- **Agent 扩展机制**：熟悉 MCP、Tool、Skill 的区别及封装调用，了解 Agent Skills 的开发与应用，能够扩展 Agent 的工具链和业务能力。
- **Prompt Engineering**：熟悉 System Prompt 设计、Few-shot 示例、CoT 思维链、JSON Mode 结构化输出等提示工程方法，能够通过提示词约束模型输出并提升稳定性。
- **RAG 检索增强生成**：掌握文档解析、Chunking、Embedding、向量化检索全流程；基于 Milvus、Chroma 等向量数据库，掌握 BM25 + 向量混合检索、RRF 融合、MMR 去重、元数据加权、动态 Top-K 与 Cross-Encoder 重排，能够优化召回率、准确率和响应时延。
- **意图识别与智能决策**：具备“规则引擎 + 小模型 + 大模型兜底”的分层意图识别经验，熟悉置信度分档、多轮上下文继承和业务隔离，能够完成规则匹配、资格判断与费用计算。
- **ChatBI 与 Text2SQL**：熟悉指标知识库、指标依赖图、层级展开、Schema Linking、表召回、字段匹配及 Join 路径规划；掌握基于 BFS 的图关系检索、Plan-and-Execute 任务拆解、Pydantic 校验、拓扑执行、SQL AST 重写与 SSE 流式输出，具备 Execution Accuracy (EA)、Exact Match (EM) 等离线评测及 Bad Case 分析经验。

工程与基础能力

- **后端与工程化能力**：熟悉 Java、Spring Boot、Python、FastAPI；熟练使用 MySQL、Redis、Docker，具备微服务接口设计与 AI 服务工程化落地能力。
- **AI 应用工程化**：具备 RAG、Agent、Function Calling 能力的服务化封装经验，熟悉 SSE 流式输出、结构化输出校验、超时重试、异常降级、调用链追踪与 Token 成本监控，能够将大模型能力稳定集成到生产业务系统。
- **Agent 扩展与开发工具**：熟悉 MCP、Tool、Skill 的区别及各自优缺点，了解 Agent Skills 的封装与调用；掌握 Codex、Cursor、Claude Code 等常见 AI 开发工具。

自我评价

- 计算机相关专业科班出身，逻辑思维清晰，具备扎实的编程基础和较强的学习能力，能够快速理解复杂业务并转化为可落地的技术方案。
- 熟练使用 Codex、Cursor 等 AI 编程工具，具备 Vibe Coding 实践，能够借助 AI 快速验证方案和原型，同时保持工程判断与结果校验。
- 具备良好的团队协作与沟通能力，能够主动同步进展和风险，与产品、研发及业务团队高效配合，推动问题闭环。