

邵景奇-AI 应用开发

邵景奇 | 男 | 22 岁

邮箱：1562142773@qq.com | 电话：17831752781

当前状态：毕业

个人简历导航助手：<https://resume.shaoji.site> (技术栈：Dify、RAG)

教育背景

浙江水利水电学院-物联网工程

2022.09 - 2026.07

专业主修：计算机网络、数据结构与算法、操作系统、数据库原理及应用、计算机组成原理等课程。成绩优异，GPA：3.44 / 5 (专业前 20%)，曾获得二、三等奖学金。

实习经历

杭州时划科技有限公司 - 智能家居事业部

2025.12 - 2026.05

智能家居设备运维与场景协同 Agent 平台

- 技术架构：
FastAPI、LangChain、FunctionCalling、MCP、MySQL、Redis、SSE
- 项目描述：
该项目面向智能家居 APP 中的设备管理、智能场景、定时任务、证书下载和后台运维场景，在现有 IoT 业务服务之上建设 Agent 协同能力，通过自然语言统一承接家庭设备查询、场景操作、定时配置、证书问题排查和设备状态分析等需求，并将用户意图转换为受权限约束的业务查询或操作任务。
- 个人职责：
 - 通过 LangChain 和 FunctionCalling 将家庭、设备、场景和消息查询能力封装为 Agent 工具，配合意图识别、上下文补全和结果汇总实现自然语言到现有业务接口的稳定调用。
 - 将 MCP 和 FunctionCalling 接入设备定时与场景定时的新增、修改和取消链路，复用分钟级对齐、预处理窗口和 Redis 分片缓存能力，并在写操作前增加参数校验和确认环节，解决自然语言指令安全落到调度系统的问题。
 - 沿用 AES、RSA 和工具返回裁剪完成证书下载问题诊断，Agent 仅接收脱敏后的设备信息和必要状态，避免 MAC 明文、证书密码和长期令牌进入模型上下文。
 - 结合 FunctionCalling 接入设备版本、连接服务和心跳统计接口，将多类后台查询结果统一整理为设备状态分析，解决运维信息分散在不同查询入口的问题。
 - 以 Redis 维护 Agent 会话上下文和授权状态，结合 AccessToken、RefreshToken 校验与工具级权限控制，实现多实例共享会话并限制 Agent 可访问的家庭、设备和场景范围。

杭州时划科技有限公司 - 智能家居事业部

2025.03 - 2026.11

金制空气 APP 功能更新

- 技术架构：
Java、Redis、AES、RSA
- 工作描述：
负责某 IoT 设备与 APP 后台平台的服务端开发，围绕“链接安全、防篡改、分钟级调度、分布式会话认证”等核心能力建设，构造安全可观测的后端服务体系。
- 主要工作：
 - 将设备 MAC 等敏感标识进行 AES 加密后参与 URL 链接拼接，防止 MAC 地址等敏感标识以明文形式出现在 URL 链接中，避免传输过程中泄露信息；对证书下载信息做 RSA 数字签名，解决了证书下载过程中下载信息被篡改的安全问题。

- 采用“分钟级对齐 + 60s 预处理窗口 + 分片缓存”的调度方案，以便执行器分钟级并发拉取，解决了触发漂移、整点雪崩与漏触发问题，显著提升设备/场景定时的稳定性与可观测性。
- 基于 Redis 的分布式 Token 管理系统，实现 AccessToken 和 RefreshToken 的双 Token 机制，AccessToken 有效期 30 分钟，RefreshToken 有效期 180 天，保证用户会话的安全性和连续性。

项目经历

知脉 Agent

2026.05-2026.08

zhimai.shaoji.site (<https://github.com/xiaoqi521521/RAG-KB>)

■ 技术架构：

RAG、FastAPI、LangChain、PostgreSQL、PGVector、SSE、MinIO、Prometheus

■ 项目描述：

基于 RAG 架构构建的企业内部知识库问答系统，支持多部门知识库隔离、文档上传与智能索引、多轮对话流式问答。用户上传企业文档（PDF / Word / Markdown / TXT），系统自动解析、分块、向量化存入 PGVector，用户提问时通过混合检索 + 精排 + 生成完成回答并标注引用来源。

■ 项目职责：

- 向量检索（PGVector 余弦相似度）与 PostgreSQL 原生全文检索并行召回，RRF 融合排序后送入 Cross-Encoder Reranker 精排，Top-K 动态调整；相比纯向量检索，Hit Rate 提升约 20%
- SSE 流式返回生成内容，每条回答标注引用来源，支持回答的幻觉检测；前端实时渲染 Markdown，流式完成后自动替换引用标记为实际文档名
- 数据权限从 Prompt 约束升级为代码硬保证，权限校验不仅在 API 层做拦截，检索阶段也做 Chunk 级别过滤，在 Chunk 检索阶段强制过滤用户无权访问的知识库的 Chunk，从数据层面杜绝跨部门信息泄露。
- Embedding、Reranker、HyDE、幻觉检测、回答生成——每个 Token 消耗环节独立埋点，按用户维度持久化到 Redis，前端监控面板实时展示各环节消耗分布
- 构建 Embedding 缓存 + 查询结果缓存 + Context Token 裁剪 三重 Token 成本控制机制，通过 Redis 复用向量化结果与高频查询结果，并按 Token 预算动态裁剪检索上下文，在保证回答质量的同时控制单次请求成本与延迟，高重复度场景的向量 Token 消耗减少约 70%，查询 Token 消耗减少约 30%
- 构建评估数据集（问题 + 期望命中 Chunk ID+ 期望答案），自动计算 Hit Rate、MRR、Faithfulness 等指标，每次调整分块策略或检索参数后跑评估对比，用数据驱动优化而非凭感觉调参

专业技能

AI 应用开发

- AI 框架与 Agent：熟练使用 Agent 开发框架（LangChain、LangGraph）；掌握 AI Agent 构建，熟悉 @Tool 工具定义、ChatMemory 多轮对话管理、流式输出及 ReAct 执行循环控制机制，掌握 Prompt 工程及优化技巧。
- RAG 检索增强：熟悉 RAG 全链路架构（文档加载、切块、向量化、检索、生成）；熟练使用混合检索、RRF 融合排序等进阶方案；能基于 PGvector 向量数据库构建 RAG 知识库，实现文档索引、向量检索和查询增强的完整流程
- AI 研发提效：熟练使用 Claude Code、Codex 等 AI 工具辅助项目搭建、代码重构及单元测试生成，并基于 SDD 范式，借助 Agent Skills 约束与引导模型输出

后端开发技术

- Java 核心：熟悉 Java 基础与集合类（如 HashMap 扩容原理）；熟练使用 JUC 并发包（线程池、锁、AQS 等）；了解 JVM 内存结构。
- 框架与微服务：熟悉 Spring、Spring Boot、MyBatis 等主流框架及核心原理（IOC、AOP、自动配置）；了解 Spring Cloud 微服务组件，理解 CAP 定理与 BASE 理论。
- 数据库与中间件：熟悉 MySQL 索引机制、事务隔离及锁；熟悉 Redis 常见缓存问题（击穿/穿透/雪崩）解决方案及持久化策略；掌握 RabbitMQ 消息可靠投递机制、死信队列及异步削峰应用。
- 多语言与扩展：熟悉 Python 语言及 FastAPI 框架；了解前端（Vue、Element UI 等）基础；了解 Linux 环境常用命令及 Git、Maven 等项目构建工具。