



河海大学
HOHAI UNIVERSITY

王圣德

电话: 18660632089 | 邮箱: 18660632089@163.com
当前状态: 在读 | 求职意向: AI Agent应用开发、AI全栈开发



教育经历

河海大学 - 电子信息类 硕士 信息科学与工程学院 全日制

2024年09月 - 2027年06月

211 双一流

河海大学 - 通信工程 本科 信息科学与工程学院 全日制

2020年09月 - 2024年06月

211 双一流

实习经历

国金证券股份有限公司 - 系统开发实习生 科技研发部 (企业技术中心)

2026年06月 - 2026年09月

- 独立负责新开户客户认领系统的全栈交付, 完成从需求 Spec、双端交互原型、React 企业微信 H5 员工/运营端及 PC 权限配置端, 到 Spring Boot + MyBatis 后端、4 张业务表、17 个 REST 接口的开发与联调; 设计 8 状态乐观锁审批状态机、角色动作下发和链式审核路由, 实现申请、审核、eBoss 审批及认领生效闭环。
- 开发 Siebel 架构逆向工程 Skill, 驱动 Agent 解析 SIF/eScript、Java、JavaScript 与接口文档, 自动梳理 Business Service 调用链、BO/BC 数据访问、业务规则及外部集成, 生成 4A 架构图、领域模型、核心流程 Mermaid 图与可追溯设计文档, 提升复杂金融遗留系统的理解与交接效率。
- 从业务沟通、流程拆解到可交互 HTML 原型开发, 支持 CRM / ISMP 复杂业务的评审与前端复用; 并采用 Vibe Coding 开发 HTML 转可编辑 Figma 工作流插件, 通过 Playwright + Scene JSON + Figma Plugin API 将页面结构、样式与交互状态自动转换为可编辑设计图层, 提升原型交付与设计迭代效率。

项目经历

PaiCLI — Java Agent CLI

2026年03月 - 至今

项目简介: 类似 Claude Code 的终端智能编程助手, 支持 ReAct 推理、Plan-and-Execute、Multi-Agent 协作、MCP 工具集成、三层记忆与 RAG 代码库检索, 可通过自然语言驱动代码开发和调试。

- 基于 ReAct 构建 Agent 核心循环, 通过 ToolRegistry 动态注册 9 个内置工具及 60+ MCP 外部工具, 由 LLM Function Calling 驱动工具选择与执行。
- 基于 ripgrep、Glob、read_file 实现 Agentic Search 精确检索, 单次搜索延迟控制在 200ms 内; 结合 RAG search_code 作为语义检索兜底, 提升代码定位与分析效率。
- 设计三层记忆系统: 短期记忆维护当前对话, 长期记忆基于 SQLite 持久化跨会话知识, 并通过 Compactor 实现边界感知的上下文压缩与 BM25 + 余弦相似度混合检索。
- 实现动态 Token 预算管理, 在上下文占用达到 90% 时自动触发摘要压缩, 控制长任务执行过程中的上下文膨胀。
- 构建 Plan-and-Execute 与 Multi-Agent 协作架构: Planner 将复杂任务拆解为 DAG, Worker 并行执行, Reviewer 审核并支持自动重试; 最大支持 4 路并行, 失败任务的下游依赖自动跳过而不阻塞独立任务。
- 设计基于 SQLite 的长任务持久化方案, 结合后台 Worker 线程池异步执行, 支持进程重启后恢复未完成任务; 集成 MCP 的 stdio / HTTP 双传输协议、Schema 自动裁剪及 HITL 审批机制, 保障工具调用安全。

派聪明RAG — 企业智能知识库问答系统

2026年02月 - 2026年04月

项目简介: 一款基于 RAG 的智能化私有知识库管理系统, 深度聚焦底层数据高精度切分、双路检索引擎设计与数据资产安全隔离, 为企业提供强上下文、快速响应的智能问答核心支撑。

- 基于 MinIO + Redis BitMap 实现分片上传与断点续传, 将 1GB 文件上传耗时从 15s 优化至 3s。
- 混合检索采用向量 KNN 召回 + 关键词/BM25 二阶段重排 (ES rescore) 的方案, 先用向量检索扩大召回窗口 (recallK=topK×30), 再在候选集上用 BM25 (AND) 精排, 并保留少量向量分作为辅助, 同时结合用户/组织/公开权限过滤, 整体相关性与稳定性提升。
- 基于 WebSocket + 大模型 Stream API 实现流式输出; Redis 存储对话历史并支持超长对话摘要记忆。
- Spring Security + JWT + RBAC 实现细粒度权限控制; 外部 API 异常时支持自动降级与兜底回复。
- 通过用户提问 + 检索片段拼接生成增强型 Prompt, 结合上下文与语义召回提升问答准确度, 构建企业私有知识问答体系。

专业技能

- Java 开发与工程能力:** 熟悉 Java 基础、集合框架、多线程、JVM 及常用设计模式; 熟悉 Spring Boot、MySQL、Redis 等 Java 后端开发技术栈, 具备 RESTful API 设计、接口开发与问题排查能力。
- 大模型基础与推理范式:** 理解大语言模型、Token、上下文窗口、Embedding 等基础概念; 熟悉 Prompt Engineering、ReAct、Plan-and-Execute、Reflection 等常见推理与 Agent 设计范式, 能够结合业务场景进行提示词优化和结构化输出设计。
- RAG 与向量检索:** 熟悉 RAG 全链路开发, 包括文档解析、文本切分、Embedding 向量化、向量检索、召回、重排序及答案生成; 了解 Milvus、pgvector 等向量数据库的基本使用与知识库问答应用构建。
- AI Agent 开发:** 熟悉 Spring AI、LangChain4j 等 Java AI 框架, 理解 Tool Calling、Function Calling、MCP、多轮对话记忆、状态管理和任务编排等核心机制; 能够开发集成外部 API、数据库或知识库的 AI Agent 应用。
- AI 辅助开发:** 熟练使用 Codex、Qoder、Claude Code 等 AI Coding 工具辅助需求分析、代码生成、调试、重构、测试及文档编写, 能够结合 AI 工具提升开发效率。
- 工程化工具:** 熟悉 Git、Maven、Docker、Linux 等常用开发与部署工具, 了解后端服务的构建、容器化部署及日志排查流程。