

李玉鑫 求职意向：开发实习生

男 | 25岁 | 手机(微信) : 18333044335 | 邮箱 : lee_yixn@163.com
现居地 : 北京市石景山区 | 状态 : 随时到岗 | 个人主页 : yixn.top



教育背景

北方工业大学 - 硕士 计算机技术

2025.09 - 2028.06

项目经验

RAG 检索增强知识库系统

2025.12 - 2026.03

项目描述：面向企业内部的 RAG 知识库系统，支持多租户文档管理，大文件分片上传与异步向量化。基于私有文档提供带引用溯源的流式问答。

技术栈：SpringBoot Kafka MySQL Redis MinIO Apache Tika Elasticsearch

- 分片上传与断点续传**：前端用 SparkMD5 指纹判重，分片后 4 路并发上传；分片存 MinIO，Redis Bitmap 记录状态，MySQL 兜底，合并时按分片序号升序拼装；中断后仅补传缺失分片，避免整体重传。
- 异步处理与可靠性**：Kafka 解耦（解析、切片、向量化、入库）全流程，上传接口不阻塞；生产端开启 acks=all + 幂等生产者 + 事务投递，将消息投递纳入本地事务规避 DB/MQ 双写不一致；消费失败固定退避重试后进入死信主题，持久化向量化状态与失败原因，避免任务静默丢失。
- 混合检索**：Elasticsearch 与 IK 分词 构建文档索引，接入 2048 维 Embedding，实现 KNN 向量召回与 BM25 重排序；融合向量余弦相似度与关键词得分，实现关键词匹配与语义相似度的混合检索。
- 实时对话链路**：基于 WebSocket 长连接与 DeepSeek StreamAPI 实现流式响应；支持中断生成并同步取消上游流；检索侧采用 Agentic RAG(ReAct)，由模型自主决定调用知识库检索，内容总结等工具进行多轮推理，基于检索片段作答并标注来源编号。

Agent CLI工具

2026.04 - 2026.08

项目描述：类 Claude Code 的终端 Coding Agent，自行实现 ReAct 循环与工具编排引擎，支持多模式执行、MCP 协议扩展、多模型适配、上下文压缩与三层安全防线。

技术栈：Java 17 JLine 4 MCP 协议 OkHttp(SSE) JGit JavaParser

- Agent 核心循环与工具编排**：实现 ReAct 主循环，模型自主决策并自决退出；多工具 4 路并发执行，结果按原始调用顺序回填保证消息协议稳定，基于工具调用签名做重复检测与轮数上限兜底；扩展 Plan-and-Execute 与 Multi-Agent 多模式执行。
- 记忆与上下文压缩**：设计对话 / 事实 / 工具结果三类记忆的注入与淘汰策略，按模型上下文窗口比例分配 Token 预算；上下文模式按窗口自适应；长期记忆结合作用域过滤，分词，时间衰减与 TopK 排序注入。
- MCP 协议工具扩展**：自研 MCP 客户端，兼容 stdio 子进程与 Streamable HTTP 两种传输；外部工具自动注册并支持列表热更新；清洗工具 schema 中的嵌套引用与超长描述，以降低模型调用失败率，环境变量延迟展开实现多源配置与故障隔离，审计日志自动脱敏。
- 三层安全防线与快照回滚**：HITL 人工审批 / 路径围栏 / 命令黑名单三层拦截，用户无法批准被策略拒绝的请求；基于 JGit 建立独立版本库快照，不污染项目 .git，支持按轮次回滚与审计脱敏。

技术技能

Agent / AI 应用：ReAct 循环与工具编排，Function Calling，Plan-and-Execute，Multi-Agent，MCP 协议，记忆体系与上下文压缩，流式与中断取消；RAG 全链路；有自主实现 Agent 引擎经验。

Agent 框架：了解 LangChain / LangGraph / LlamaIndex，(LCEL编排，Tool/Retriever/Memory抽象，StateGraph状态图，Human-in-the-Loop)；了解Spring AI, LangChain4j.

编程语言：Java(集合，并发，JVM 基础); Python(asyncio, Pydantic, FastAPI).

数据与中间件：MySQL(索引，事务，锁); Redis(缓存三大问题，持久化，Bitmap); Kafka(acks，幂等生产者，死信); Elasticsearch(倒排索引，中文分词，向量检索).

开发框架：Spring Boot, Spring Security, Spring Data JPA; 了解IoC / AOP基础原理。

工程与工具：Git, Maven, Docker, Nginx, Linux, Python / Shell 脚本。

其他：英语六级CET-6；熟练使用 Cursor, Codex, DeepSeek Harness等 AI 编程工具辅助开发与调试。