

孙建铭

AI Agent 开发工程师 / RAG 与大模型应用开发

意向城市 西安 · 本科 · 16623875909

16623875909@163.com / 在线作品 ownmind.site

Python / FastAPI · 混合检索 · 工具调用 · 应用交付

01 个人简介

PROFILE

现负责企业内部 RAG 知识问答，将技术资料、故障排查记录与 FAQ 接入受权限和版本约束的混合检索链路，并提供来源引用。个人独立交付孙郎 App，将多轮对话、私有知识和确认后写入接入 Android / Web，覆盖后端开发、回归验证与云端部署。

02 专业技能

SKILLS

Agent 编排 个人求职 Agent 基于 Python、FastAPI 与 LangGraph 实现多轮对话、检索编排和受限只读工具；孙郎 App 将业务写入拆为模型草稿、用户确认和幂等事务提交。

RAG 检索 具备文档解析、切块、Embedding、**BM25**、向量召回、**RRF** 和来源引用实践；孙郎 App 使用 pgvector 做用户与 **READY** 状态过滤，个人求职 Agent 使用 Qdrant 做版本化向量索引。

检索增强 个人求职 Agent 使用 BGE 中文向量、交叉编码器重排和 Redis 分数缓存；为向量检索与重排异常设计可观测的降级路径。

多轮与记忆 孙郎 App 支持多轮主题承接、会话摘要和 Android 本机偏好记忆；通过 Keystore AES-GCM 按账号加密保存，支持修改、删除与关闭。

工程交付 个人项目使用 React、PostgreSQL、异步 Worker、Docker Compose、Linux/Nginx；完成 Android/Web 联调、回归验证、签名 APK 与备份恢复。

03 工作经历

EXPERIENCE

南京瑞辉数据有限公司

2026.02 - 至今

企业知识问答 负责技术资料、故障排查记录与 FAQ 的接入和检索链路，维护资料来源、可见范围及有效版本，为带出处回答提供可追溯证据。

检索与验证 实现 **BM25** 与向量双路召回及 **RRF** 融合，在候选范围中加入权限和有效版本约束；以多来源、旧版排除和空证据场景检查检索及引用结果。

工程经历 具备相机启动链路与 HAL Trace 性能验证经验，利用日志和启动时序数据支持问题定位及版本回归。

重庆智呈映趣有限公司

2025.07 - 2026.02

Camera 影像开发 参与 Android 15 / 16 相机 App 适配，使用 Logcat、ANR 与 Perfetto 定位兼容性和性能问题，完成修复与版本回归。

04 核心项目

SELECTED PROJECTS

企业内部 RAG 知识问答系统

企业项目 | 2026.02 - 至今

- 链路设计** 资料接入后按来源、有效版本和可见范围限定候选；经 **BM25** 与向量双路召回、**RRF** 融合去重后，组织带出处的回答上下文。
- 检索取舍** **BM25** 用于错误码、专业术语与关键词匹配，向量召回补充语义相近候选；**RRF** 基于排名融合，避免直接相加不同尺度分数。
- 范围控制** 按文档与章节维护来源、版本和可见范围，避免旧版本或无权限内容进入模型上下文。
- 回答验证** 回答附资料出处；无适用证据时返回证据不足提示，以多来源、旧版排除、权限和空证据场景检查检索及引用。

孙郎 AI 生活助手

个人项目 | 独立主导 | 2026.07 - 至今

- 完整交付** 独立完成从需求梳理到 Android/Web 联调、云端部署和签名 APK 交付，提供多轮对话、私有知识问答、待办与记账，并开展小范围试用。
- 私有知识** 搭建解析、切块、Embedding 与检索链路；仅检索当前用户可见且已完成索引的 **READY** 资料，保留回答来源元数据。
- 工具写入** 将操作拆为模型草稿、用户编辑、明确确认和事务提交；服务端重新校验权限，使用**幂等键**防止重复写入待办或账目。
- 记忆与语音** 通过 Android Keystore AES-GCM 按账号加密保存回复偏好；SenseVoice Small INT8 本机转写失败或超时后回退服务器，支持修正与重试。
- 验证与发布** 在真实数据库副本中验证并发与重复草稿确认下只产生一次业务写入；完成业务后端回归、数据迁移、备份恢复和镜像回退验证。

05 教育背景

EDUCATION

重庆财经学院

2021.09 - 2025.06

智能科学与技术 · 本科

06 荣誉奖项

AWARDS

- 中国软件杯** 大学生软件设计大赛 全国总决赛三等奖
- 计算机设计大赛** 中国大学生计算机设计大赛 全国赛三等奖