

孙堂铭

邮箱: 516990314@qq.com

电话: 17779711456

Java全栈/AI应用开发

地址: 江西赣州

个人网站: stm2026.pages.dev

教育经历

赣南师范大学

计算机科学与技术

2023/09 - 2027/06

荣誉证书: 获评校一、二等奖学金（专业前5%），获校“三好学生”等荣誉称号，软件设计师（中级），通过CET-4

竞赛结果: 获江西省数学建模本科生组大赛一等奖，校ICPC竞赛二等奖，中国机器人及人工智能大赛省三

校园经历: 担任校研究中心/实验室的技术负责人，带领团队完成开发需求及网站维护，有较强的抗压能力

工作经验

珠海鑫链科技有限公司

全栈开发实习生

2026/06 - 至今

QueryMind 自然语言查数平台（NL2SQL） | 全栈开发

前端 Vue3 | 后端 Go/Gin | AI 引擎 Python/FastAPI + LlamaIndex | Milvus · MySQL/PostgreSQL · Redis

多租户 NL2SQL 查数平台: 中文提问 → SQL 生成 → 安全执行 → 多形态展示（图表 / 表格 / SQL / 中文释义），业务人员零代码自助取数，覆盖 7 个数据源、333 张业务表;独立完成四层架构全栈开发

攻克「大 Schema 撑爆 Prompt」难题: 全量 DDL 约 6 万 token 远超本地模型 8K 上下文窗口，设计 **RAG 多路召回**（Milvus 向量 + BM25 词法 + Few-shot 反哺）按需动态召回 Top-K 表 DDL，轻量上下文稳定生成多表 JOIN 查询

构建 SQL 生成安全链路: 仅允许 SELECT，注入检测（注释符/UNION 拦截），语法/笛卡尔自检拦截不合格 SQL，校验失败将错误回传模型至多一次自纠，全链路 token 计量，延迟与成本可控

落地多租户权限安全体系: 三级角色 + API/表/字段级权限，端到端 role_id 权限链杜绝绕过引擎直调；召回侧按角色允许表集合事前过滤 + 生成 SQL 经 sqlglot 表级/列级硬闸门事后校验，高危写操作 100% 拦截、**审计全链路留痕**

搭建 RAGAS 离线评测体系: 自研 SQL 接地校验 + LLM 语义判分替代指标（原生指标对 SQL 失真），**Langfuse** 全流程追踪支撑 badcase 归因，驱动召回 top-k、样本等参数迭代调优，评测集语法正确率 97%、执行准确率 87.9%

绍兴易捷信息技术有限公司

Java全栈开发工程师

2025/6~2025/10

主要职责: 参与公司的WMS仓储管理系统和嵊州市教育平台的需求开发及维护

1、**第三方系统对接:** 使用适配器模式封装供应商API，并全程负责接口开发、联调、测试与线上问题跟踪，确保项目落地。实现业务代码与第三方解耦，新供应商接入无需改动核心逻辑。

2、设计并实现**大文件导入导出系统**，通过文件分片、线程池异步处理和断点续传机制，有效解决内存溢出与超时问题，并引入任务队列管理实现资源可控的全局下载，保障系统处理数据的稳定性与用户体验

3、**信创数据库国产化迁移:** 负责以“业务代码零修改”为目标，通过设计与改造数据访问中间件（SQL拦截器），完成**MySQL至达梦8的无感迁移与全量数据同步**，覆盖95%以上常用CRUD场景

4、实现**定制化文档生成功能**，基于业务需求开发Excel、Word、PDF等文件的动态模板，支持页眉、封面、合并单元格等**复杂格式定制**，满足多场景下的文档自动化输出需求

项目经历

“悦题”-判题服务系统

独立全栈开发者 & 项目负责人

2025/03 - 2025/05

<https://gitee.com/hutao511/JoyJudge>

项目简介: “悦题”是一个基于微服务架构的智能教育平台，基于SpringCloud + Vue3框架搭建。它以在线判题与AI解题为核心，集成作业管理、班级协作、多模态作答、实时协同、错题巩固与积分激励等功能

关键功能实现:

· **数据能力扩展:** 基于 FastAPI 独立构建 Python 计算层，经 MCP 协议对接 Java 主系统，将数据清洗、数值计算、Matplotlib 渲染等计算密集型任务卸载至独立算力节点，与 AI 推理异步并行，避免计算任务阻塞主服务

· **AI解答:** 搭建 RAG 问答链路（文档解析、分块入库、ES 关键词 + 语义双路召回、RRF 融合重排、置信过滤去重）抑制幻觉；基于 ReAct 构建答疑 Agent，并行调度 Tavily、错题检索，渐进式披露关键证据、降低 Token 消耗

· **检索与缓存优化:** 基于 Elasticsearch 构建全文检索提升题目查找效率；前端 Pinia + 后端 Caffeine + Redis 哨兵集群三级缓存热门题目，查询响应从 200ms 降至 50ms

· **接口限流:** 基于 Redisson+AOP 实现用户级令牌桶限流，结合 Sentinel 完成集群级流量控制，支持 Nacos 动态配置，**有效防护恶意请求**，保障核心判题服务高可用

专业技能

AI 应用: 熟练使用 LangChain、SpringAI、LlamaIndex等AI应用框架完成 RAG 知识库问答与智能体应用开发

后端开发: 掌握 Java 语法与集合、并发基础，熟悉 Spring 生态，熟悉基于 Python / Go 的 Web 框架开发

数据存储与中间件: 熟悉 MySQL、PostgreSQL的使用与核心原理，熟悉 Redis、RabbitMQ的典型应用场景

全栈开发: 熟练使用 Vue、Uniapp、Electron，具备需求分析、设计开发测试、部署运维迭代的全流程交付能力

开发工具链: 熟练使用Git、Maven、Docker、IDEA、Apifox、Cursor、Codex、影刀RPA等工具