



孙堂铭

AI全栈开发

邮箱: 516990314@qq.com

地址: 江西赣州

电话: 17779711456

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

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: 熟悉 Java 语法与集合、并发基础, 熟悉 Spring 生态, 熟悉 Go + Gin、Python + FastAPI 的 Web 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

工作经验

珠海鑫链科技有限公司 全栈开发实习生 2026/06 - 2026/09

QueryMind 自然语言查数平台(NL2SQL) | 独立开发·从0到1

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, 已交付内部测试使用

搭建RAG检索召回方案: 全量schema约6万token远超本地模型8K上下文窗口, 基于 Milvus 使用 HNSW_SQ(SQ8) 索引, 实现稠密-稀疏向量混合检索 + Few-shot 反哺, 按需 Top-K 召回表结构, 轻量上下文稳定生成多表 JOIN 查询

构建 SQL 生成安全执行链路: 通过 AST 语法白名单与数据库只读账号强制仅 SELECT 查询, 覆盖注入特征拦截、语法校验与笛卡尔积静态检测, 校验失败触发至多一次模型自纠, 配套全链路 token 计量限额, 保障延迟与成本可控

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%

机场快线综合运营平台(客运行业一体化 SaaS) | 带领3人团队分领域模块开发

后端 Go/Gin | 消息队列 RabbitMQ | 支付 微信支付 APIv3 | 前端 Android(Kotlin) · uni-app · Vue3

项目背景: 面向乘客、司机与运营三端全链路, 乘客微信小程序实名购票 → 运营web后台调度派车 → 司机移动端站点打卡与验票履约 → 异常售后 → 财务对账结算 → 监管报送, 后端包含8大业务域、122张表、318个测试文件

区间容量扣减重构: 一张票跨多段运力, 须逐段扣减取锁, 数据库往返随区间数线性增长; 扣减判定须与台账写入同一事务, 故未采用 Redis 预扣, 改为单事务内一条批量更新扣完全区间并锁行、批量写台账流水与占用; 更新影响行数不足区间数即整体回滚。数据库往返由 3N 降至 3, E2E 验证中余额为 1 时多协程并发下单仅一单成功

退款资金一致性闭环: 车票作废、资金冲回、运力回补三类动作收敛于同一事务, 任一失败整体回滚, 不留“钱已退、票仍可验”的半截状态; 退款 5 态流转, 账务只追加、不修改, 冲正分录等额反向冲回; 外部渠道无法进本地事务, 失败按差错分级转人工工单; 审计留痕独立落库、不随回滚消失。失败注入验证通过, 冲正分录逐笔可核对

跨端幂等键策略: 断网重传、双击、超时重试均命中同一把幂等键。同键同内容重放首次响应、异内容返回状态码 422; 在线键只存客户端内存; 离线补传按(设备, 序号)确定性派生、重启仍同一笔; 附件键按内容哈希派生; Redis SETNX 前置拦截、MySQL 唯一键兜底。各端可安全重试, 三类重复场景经用例验证全部被拦截、未产生第二笔

监管报送可靠投递: 报送失败不回滚已成交业务、不漏报。MQ 只能保证送入 broker、无法与业务库同事务, 改用本地消息表: 出票事务内只落一条 PENDING 记录, 多实例异步补发、限时抢占投递权以免并发重投; 重试耗尽转死信告警; 日终按出票数与上报记录数核对(不用成功数, 避免误报)。故障演练验证, 投递可自动恢复、核对一致

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、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
- 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间问题，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、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。
- 4、实现定制化文档生成功能，基于 Apache POI 完成 Excel、Word 动态模板开发，支持页眉、封面、合并单元格等复杂格式定制；集成 docx4j 转 PDF 并对接 kkFileView 在线双端预览，满足多场景下的文档自动化输出需求

项目经历

“悦题”-判题服务平台

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。获校原创软件大赛一等奖

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，可并行调度 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：针对题库检索慢、热题重复查询压库问题，基于 Elasticsearch 实现全文检索，搭建前端 Pinia、服务端 Caffeine、Redis 哨兵集群三级缓存体系，查询响应从 200ms 降至 50ms
- 接口限流防护：针对作业高峰流量突增、恶意刷取请求风险，基于 Redisson + AOP 实现用户级令牌桶限流，结合 Sentinel 做集群流量管控与 OpenFeign 调用兜底，支持 Nacos 动态配置，保障核心判题服务高可用