

蒋文睿

2002-04 (24 岁) · 籍贯：广西南宁
手机微信同号：17852636053 · 求职意向：Agent/Java 开发
邮箱：jiang.wenrui@outlook.com
个人主页 (更多联系方式)：https://www.jiangwenrui.com



教育背景

北京交通大学 (211 本科) - 计算机科学与技术 2020.09-2024.06
First class honor - 学业三等奖学金 - GPA: 3.73/4.00
北京交通大学 (211 硕士) - 计算机科学与技术 - 研究生在读 2024.09-2027.06
一等奖学金

专业技能

计算机基础：操作系统、计算机网络、数据结构与算法等。
Java 后端：熟练掌握 Java 基础、集合、并发、JVM 基础、Spring Boot、Spring Cloud、MyBatis-Plus、Maven 等技术栈。
Agent 应用：熟悉 Context/Harness Engineering、RAG、MCP、Tool Calling、LangGraph 工作流编排及 ReAct、Plan-and-Execute 等 Agent 架构设计。
中间件与工程：熟悉 Git、MySQL、Redis、MongoDB、Elasticsearch、Kafka、RabbitMQ、Nginx、Docker、Neo4j、MinIO、XXL-JOB 等组件。
线上运维与分布式：熟悉 Linux、日志分析与 CI/CD 流程，了解服务监控、链路追踪、幂等设计、分布式事务、数据一致性、限流、熔断及重试机制。

项目经历

医疗信息集成平台导诊 Agent 接入-后端/Agent 应用开发 2026.06-2026.07

项目地址：https://medical.jiangwenrui.com/

项目简介：面向城市级医疗信息整合平台，平台汇聚当地医疗机构的科室、医生及排班等信息，并基于 Spring Cloud 微服务架构提供医疗预约服务。在原有医院、用户、订单等核心服务解耦运行的情况下，新增智能导诊 Agent 微服务，通过对接既有业务接口，为患者提供症状理解、就诊意图识别、科室推荐、号源查询、挂号确认及预诊报告生成等能力，降低患者理解医院科室体系和选择就诊科室的门槛，解决挂号意图模糊、科室匹配困难等问题。

项目技术栈：Java、Spring Boot、Spring Cloud、OpenFeign、DeepSeek API、Function Calling / Tool Calling、Agent Workflow、Session State、Human-in-the-loop、Redis、MySQL、MongoDB、JWT、Nginx、Docker
我的工作：

- 设计独立 Agent 微服务接入方案，将导诊编排与医院、用户、订单等核心服务解耦，通过 OpenFeign 封装 8 类白名单工具，并基于 DeepSeek 原生 Tool Calling 实现多轮“模型决策—工具执行—结果回传”闭环。
- 设计症状收集、科室推荐、号源查询、挂号确认、订单提交及就诊引导状态机，通过 slots 维护多轮上下文，并结合急症规则、否定症状识别及 Human-in-the-loop 确认机制阻止模型越权下单。
- 构建覆盖急症识别、否定症状、工具路由、参数完整性及下单保护的 Agent 评测集，在 XX 条测试用例上任务完成率达到 XX%、工具调用准确率 XX%，P95 响应延迟为 XX ms。
- 基于 Redis + PostgreSQL/pgvector 构建短期会话与长期记忆双层存储，实现多实例会话共享、服务重启恢复及用户级记忆隔离，Recall@5 达到 XX%、P95 检索延迟为 XX ms。
- 基于 Redis Lua 原子令牌与订单业务唯一约束实现挂号接口幂等，并为 OpenFeign 工具调用增加超时、有限重试、熔断和降级，在 XX 次重复提交压测中重复订单数为 0。
- 构建面向医院科室、挂号规则及就诊须知的 RAG 知识库，实现文档解析、混合检索、引用溯源和增量更新，Recall@5 达到 XX%、引用准确率达到 XX%。

实习经历

北京北九方轨道交通科技有限公司 - 后端项目实习生

2025.04-2025.07

项目简介：面向某大型重载铁路专线车辆段的机车检修场景，针对传统人工提票效率低、信息汇总慢、故障类型及车辆病害分布难以追踪等问题，对检修提票流程进行数字化改造，实现提票录入与检索、检修文档及图片上传、算法识别结果回写和业务数据汇总、统一调度检修，为行车修理全生命周期实现数字化转型。

我的工作：

- 负责机车检修提票核心模块，设计提票、附件、故障信息及识别结果等数据模型，并通过状态机管理待提交、待识别、待复核和已完成等业务状态，支撑检修流程全过程追踪。
- 基于 Spring Security、JWT 和 RBAC 实现统一认证鉴权，设计用户—角色—权限关联模型，并结合车辆段、部门和岗位限制数据访问范围，避免普通检修人员越权查看或修改工单。
- 设计检修文档与图片处理链路，使用 MinIO 存储文件、MySQL 保存附件元数据，通过 RabbitMQ 异步触发算法识别；对结果回调增加业务编号、状态校验和幂等控制，支持失败重试及人工复核。
- 针对多条件组合检索和故障统计场景优化 MySQL 索引与分页查询，支持按车型、部件、故障类型及时间范围汇总检修数据；经构造 XX 万条工单测试，P95 查询耗时由 XX ms 降至 XX ms。

北京合力云通科技有限公司 - Java 开发工程师（实习）

2022.12-2023.03

项目简介：A/B 测试平台用于对网页或应用的多个版本进行实验分析，根据用户行为数据和算法结果确定更优版本，并动态调整流量分配比例。

项目架构：Spring Cloud、MyBatis、MySQL、RabbitMQ、XXL-JOB

- 负责 A/B 实验生命周期与调度模块，设计草稿、待启动、运行中、暂停及已结束等实验状态，基于 XXL-JOB 按实验时间窗口完成定时启停，并通过版本号校验和接口幂等处理重复回调及配置并发修改。
- 设计用户稳定分流方案，基于实验编号与用户标识进行哈希分桶，保证同一用户在实验周期内命中固定版本；通过保存实验配置快照隔离不同版本数据，避免中途修改配置污染历史实验结果。
- 基于 RabbitMQ 异步接收曝光、点击和转化事件，使用唯一约束实现消费幂等，按实验和版本聚合 UV、CTR、CVR 等指标；经压测达到 XX TPS，重复事件过滤率达到 XX%。
- 将聚合结果接入 Epsilon-Greedy、UCB1 等多臂老虎机策略，通过最小样本量、单次最大调流幅度和异常指标回滚机制限制自动调流风险，并生成实验报告发送至客户邮箱。

科研成果、英语能力

- Causal Grouping Network for News-driven Stock Forecasting 一作（KDD 在审）
- CET-4: 542
- CET-6: 539

自我评价

我责任心强、执行力好，重视人际交往，拥有国际视野，重视团队协作，热爱并善于钻研新技术新工具，具有良好的自驱力和学习能力，能够在复杂任务及高压环境中保持稳定推进。