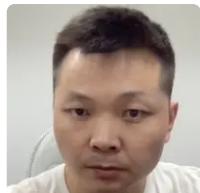


刘浩

男 | 年龄：38岁 | ☎ 13564496085 | 📩 593954214@qq.com

14年工作经验 | 求职意向：自然语言处理算法 | 期望薪资：45-60K | 期望城市：上海



个人优势

拥有14年金融IT经验，精通Java后端开发，5年大语言模型（LLM）及AI项目实战经验。擅长微服务架构、系统改造、系统调优及容器化部署，熟练掌握Spring全家桶、Docker、Kubernetes、OpenShift等技术。

专注NLP、AI Agent, AIGC与LLM（Gemini、Azure OpenAI、LLaMA）企业落地，精通PyTorch/TensorFlow，熟练运用RAG、MCP、LoRA微调与强化学习。曾主导多个AI产品成功上线，如Smart Jira、智能助手等，显著提升业务效率。

英语流利，具备跨职能团队领导力与强执行力，擅长从零高效交付创新型AI项目。

个人CSDN博客访问量超百万（1,062,343+），原创技术文章302篇，链接：<https://harryliu.blog.csdn.net/>

工作经历

花旗金融信息服务（中国）有限公司 大模型算法

2014.06-至今

内容：

- 拥有扎实的Java Web后端开发经验，近年来聚焦AI技术方向，深入自然语言处理（NLP）、深度学习及大语言模型（LLM）应用，成功主导多个企业级AI项目的研发与落地。
- 担任MOT部门AI兴趣小组组长兼讲师（2023-2025），带领团队从零起步掌握大模型相关技术，推动Smart Jira、BizProb Bot、Smart Email Finder等创新项目成功上线，并多次获得部门级及公司级创新奖项。
- 在实际应用中，基于Gemini、Azure OpenAI、LLaMA3等平台，开发多项智能系统，如文本自动格式化工具、文档智能重写模型等，有效提升运营效率与数据质量。熟练掌握RAG、LangChain、FastAPI等技术栈，构建企业级LLM服务，加速传统业务的智能化升级。

业绩：

- 带领MOT团队完成AI能力从0到1的建设，成功上线多个GenAI项目
- 2024年主导“Smart Jira”，获公司GenAI创新大赛一等奖（奖金¥15,000）
- 2022年“Smart Email Finder”获创新达人赛A/B轮奖金共¥12,000
- 2021年开发Raspberry Pi AI项目“123木头人”，获Unit2一等奖
- 2010年设计邮件自动回复系统，获创新达人赛A/B轮共¥12,000奖金

花旗金融信息服务(中国)有限公司 Java

2014.01-至今

内容：

“核心金融系统微服务化与容器化改造”

主导将核心金融系统多个服务从WebLogic平滑迁移至Spring Boot架构，采用Netflix技术栈（Eureka、Feign、Zuul）构建弹性服务治理体系。带领团队完成数十个微服务的Docker化及OpenShift私有云部署，实现零停机混合部署。

负责系统升级，包括Spring Boot 2到3、Hibernate、Spring Cloud、JDK 17升级，消息中间件由Tibco Queue替换为Kafka，通信协议由Hessian改为REST API。设计并交付基于FastAPI的Oracle数据归档系统，优化数据库性能。构建Watch Tower容器可视化平台，实现多环境监控与自动弹性伸缩。

业绩：

1. 2024 年，4 个月内完成 100+ 代码库升级至 Spring Boot 3.2.4，推动 JDK 17 全面适配
2. 2023 年，主导 50+ 微服务迁移至 OpenShift，构建自动弹性伸缩机制
3. 2017 年，独立交付 Swap 项目，实现零 Bug 生产上线，获项目经理表彰

文思海辉技术有限公司 Java 2012.07-2014.06

外派至花旗中国，负责企业级金融系统的维护与迭代开发。使用Java与Spring, hibernate 框架开发后端模块，配合前端 GWT(Google)实现交互功能。参与需求分析、系统设计、单元测试与部署流程，提升系统稳定性与用户体验。表现优异，于2014年成功转为花旗正式员工。

印孚瑟斯技术（中国）有限公司 Java 2011.06-2012.06

参与为大众汽车开发的供应商评分系统的全流程开发。负责Java后端接口开发及前端jQuery页面交互设计。熟悉软件开发流程及国际化项目协作标准，具备扎实编码能力和文档撰写能力。

项目经历

[AI-LLM] 项目名称：BizPro Bot（业务问题智能问答助手） 项目负责人 2025.01-至今

基于 Gemini-1.5-Pro 构建的企业智能助手，提升整个部门中员工获取业务信息的效率与准确性。核心功能包括：

1. 文档问答与摘要：支持上传内部文档，自动生成摘要并智能回答相关问题；
2. 系统业务知识助手：根据用户的问题与图片，实时抓取相关JIRA数据，构建问答上下文，提供精准答复。

项目采用 RAG 框架、Hybrid Search (MD25 + 向量检索)、Cross-Encoder 重排序等技术，解决信息碎片化与搜索低效问题，显著提升运营与支持效率。

[AI-LLM] 自动代码审查ACR 项目负责人 2024.06-至今

设计并开发企业内部代码审查平台 ACR，集成至 Bitbucket PR 流程，基于 Gemini / GPT-4 等大语言模型自动生成结构化审查建议，降低人工 review 成本、提升代码质量。

主要功能包括：自动抓取代码变更、生成智能审查建议、自动提交 PR 评论，并通过 Webhook 实现实时触发与反馈。

[AI-LLM] 微调的业务问答Bot 项目负责人 2024.03-至今

基于企业内部 JIRA 数据，构建智能业务问答机器人。通过 LLaMA 3-8B 模型结合 LoRA 微调和 DPO 强化学习，赋予模型业务知识并优化交互体验。集成 RAG 技术实现向量检索 + 生成式回答，提升复杂问题处理能力；采用增量训练机制持续更新知识，管理多阶段适配器版本，保障训练稳定性。系统训练与部署流程高度自动化，支持端到端高效运行。

[AI-LLM] Smart JIRA 智能工单生成工具 项目负责人 2024.06-2025.02

主导开发 Smart Jira，基于 Gemini-1.5 Pro / 2.0 Flash 自动解析文本、图片、邮件、语音等需求信息，生成结构化 Jira 工单内容 (Summary、Description、Acceptance Criteria、Story Points、Issue Type 等)，显著提升项目管理效率。已在 MOT 部门上线，正向全公司推广。

亮点包括：多格式输入支持、规范工单生成、自动创建工单、减少人工错误、加速敏捷流程。

[AI-LLM] 文本自动更新 项目负责人 2024.06-2025.01

该系统自动处理和更新上游传递的事件描述文本，取代了传统人工修改全大写文本的低效流程。项目利用 Python difflib 精准提取文本差异，将差异和之前的格式化文本输入 GenAI (如 Gemini 或 Azure OpenAI)，自动生成更新后的客户文本。实现全自动化处理，大幅提升文本一致性和交付效率，已成功应用于日常事件处理，显著提升客户通信自动化水平。

基于生成式 AI，自动对企业行动中每个叙述性文本（narrative text）做文本格式好，包括大写转换合适大小写、拼写、语法、格式与缩略词规范化，已集成至 ASPEN 平台。

项目成效：日均处理 35 条文本，单条节省约 19 分钟，显著提升文本规范性与一致性，降低人工成本，并为后续流程智能化提供模板。

[Java] Springboot2.x 到Springboot 3.x升级和JDK17 升级 项目负责人

2024.06-2024.11

带领团队在不中断业务的前提下，完成 ASPEN 系统中 100+ 代码库（含 50+ 微服务与公共 Jar 包）的升级改造工作。主要内容包括：

1. 将 Spring Boot 升级至 3.2.4, Hibernate 升级至 6.x, Spring Cloud 升级至 2023.0.5;
2. 替换 Eureka、Spring Cloud Gateway、OpenFeign 等核心组件；
3. 从 JDK 8 升级到 JDK 17，解决兼容性问题。
4. 负责升级方案设计、依赖适配与兼容性分析，确保系统平稳过渡。

[Java] Spring Cloud Config 集成与配置中心重构 项目负责人

2024.02-2024.11

主导引入 Spring Cloud Config 架构，替代原有依赖独立 config-content.jar 的配置方式，实现配置集中管理与动态热更新。后端使用 Oracle 存储配置信息，通过集成 Kafka 实现配置变更通知，无需重启微服务即可实时刷新配置。

负责整体架构设计、Config Server 搭建与微服务端改造，制定统一配置模板与上线流程，保障100+个微服务平稳迁移。该项目显著提升了配置变更效率与系统可维护性，降低了人为误操作风险，运维效率提升超50%。

[Java] Watch Tower 项目负责人

2024.01-2024.07

为提升 Dev、UAT、PROD 三个环境中多个 Red Hat OpenShift 容器平台的运维效率，主导开发 Watch Tower 平台，实现对各环境中所有微服务的集中可视化管理。

通过构建统一的 Dashboard，展示每个服务所在的 OpenShift 集群、当前实例数量（Pods）、部署状态等关键信息，支持一键操作如 Scale Up/Down、服务定位等，显著简化运维流程。

我负责系统设计、数据采集逻辑开发、界面构建与后端服务编排，确保平台性能稳定、操作直观，帮助团队从原本分散的命令行查询方式转向集中式可视化运维，大幅提升运维响应速度与准确性。

[Java-Cloud] DevStream-自动化 CI/CD 构建与部署平台 框架设计

2023.05-2024.06

基于Bitbucket中的Webhook 驱动，为 Bitbucket 项目提供自动化构建与部署能力。开发者提交代码后，系统自动触发 Jenkins Pipeline，依次完成 Maven 构建、Sonar 代码检查、Docker 镜像构建及 UDeploy 部署。

所有流程通过统一的 Groovy 脚本实现，脚本集中管理、动态调用，确保高可维护性和复用性。该系统显著提升了交付效率，降低人工成本，适用于大规模微服务的持续集成与快速部署场景。

[Java-Cloud] 微服务上私有云的迁移 项目负责人

2023.06-2024.05

带领5人团队完成企业级微服务从传统 Linux/虚拟机部署向 OpenShift 私有云的全面迁移。原系统包含多个基于 VM/Linux 的部署环境，迁移目标为将每个微服务容器化，采用 LightSpeed 执行 pipeline (代码安全检查，Maven Building, Docker Image Building, Udeploy 部署)，并部署至 OpenShift 平台，确保服务可独立运行。

[Java] Tibco Queue 到 Kafka & Hessian 到 REST APII 迁移改造 项目负责人

2023.07-2024.02

带领4人团队完成两个关键通信架构的升级改造工作。在保证业务连续性和版本迭代节奏的前提下，顺利将微服务间的异步通信机制从 Tibco Queue 平稳迁移至 Kafka，提高系统吞吐能力和消息可观测性。

同步完成微服务间同步通信方式从 Hessian 调用迁移为标准的 Spring REST API，解决 Hessian 在 JDK 1.7 环境下因 Jakarta EE 标准兼容性问题而无法使用的技术瓶颈。

OpenShift Pod 自动弹性伸缩系统 项目负责人

2023.02-2023.05

基于实时监控指标（包括 Tibco 队列大小、Kafka 消息数量、内存使用率、CPU 利用率及时间段），开发自动弹性伸缩解决方案。系统通过调用 OpenShift CLI 命令及 Kubernetes SDK，动态调整特定微服务的 Pod 数量，实现根据业务负载自动扩容或缩容。

该方案有效提升了资源利用率，保障系统高可用性与响应性能，降低了人工运维干预，实现微服务的智能弹性管理。

[Java-Micro] 微服务架构重构与 Netflix 技术栈引入 项目负责人

2018.02-2019.03

主导将部署在 WebLogic 的多个服务重构为基于 Spring Boot 的独立微服务，提升系统可扩展性与故障隔离能力。引入 Netflix 技术栈（Eureka、Feign、Zuul）实现服务注册、调用与统一网关，优化系统架构，实现自动注册、动态路由与松耦合通信，显著提升灵活性与部署效率。

教育经历

复旦大学 硕士 软件工程（人工智能方向）

2019-2022

延安大学 本科 计算机科学与技术

2007-2011