



刘浩

在职，急寻新工作 · 37岁 · 硕士 · 工作14年2个月

13564496085

593954214@qq.com

myhomestar

优势亮点

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

2019/06-至今

VP-C13-自然语言处理（NLP）

- 近年专注于AI方向，深入NLP、深度学习与大语言模型（LLM）应用，主导多个企业级AI项目成功落地。
- 作为MOT部门AI兴趣小组组长兼讲师（2023–2025），带领团队从零构建大模型能力，推动Smart Jira、BizProb Bot、Smart Email Finder等创新项目上线，屡获公司级创新奖。
- 基于Gemini、Azure OpenAI、LLaMA3等平台，自动化文本处理任务，显著提升运营效率与数据质量。熟练掌握RAG、AI Agent、MCP、LoRA等技术，构建企业级LLM服务体系，赋能业务智能化转型。

代表业绩：

2. 2024 年主导 “Smart Jira”，获MOT部门 GenAI 创新大赛一等奖（奖金 ¥15,000）
3. 2022 年 “Smart Email Finder” 获创新达人赛 A/B 轮奖金共 ¥12,000
4. 2021 年开发 Raspberry Pi AI 项目“123 木头人”，获 Unit2 一等奖
5. 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

2014/06-2019/06

VP-C13-Java高级软件工程师

1. 主导核心服务从 WebLogic 平滑迁移至 Spring Boot 架构，构建基于 Eureka、Feign、Zuul 的弹性服务治理体系。带领团队完成数十个微服务的 Docker 化及 OpenShift 私有云部署，实现零停机混合上线。
2. 推动技术栈全面升级（Spring Boot 2→3、Hibernate、Spring Cloud、JDK 17），将消息中间件由 Tibco 替换为 Kafka，通信协议由 Hessian 迁移至 REST，显著提升系统稳定性与可维护性。
3. 优化系统与批处理性能，完成 Oracle 历史数据归档，提升整体运行效率。自主建设 Watch Tower 容器监控平台，实现多环境可视化管理与自动弹性伸缩能力。

关键业绩：

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

2012/07-2014/06

Java

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。表现优异，于2014年成功转为花旗正式员工。

Infosys

2011/07-2012/07

Java

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。

项目经验

[AI-LLM] BizPro Bot 企业级智能问答助手

2025/01-至今

项目负责人

项目描述

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] OCQuery Agent(基于自然语言的 OpenShift 控制 AI Agent)

2024/12-至今

项目负责人

项目描述

构建了一个基于大语言模型的 AI Agent —— OCQuery，支持通过自然语言控制 OpenShift 集群。用户只需输入指令（如“查看某个服务的异常日志”），Agent 即可自动生成并执行对应的 Python 脚本和 oc 命令，完成资源查询、日志分析等任务，并将结果智能解释返回。项目融合了 GPT-4、LangChain、沙箱代码执行器、OpenShift CLI 等关键技术，有效提升了运维操作的自动化与交互体验，同时确保命令执行的可控性与安全性。

[AI-LLM] ACR 自动代码审查平台

2024/06-至今

项目负责人

项目描述

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] Smart JIRA 智能工单生成工具2024/06-至今

项目负责人

项目描述

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] 微调的业务问答Bot2024/03-至今

项目负责人

项目描述

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] 文本自动更新2024/06-2025/04

项目负责人

项目描述

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。

[AI-LLM] 自动文本格式化2024/01-2025/02

项目负责人

项目描述

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。

[AI-Deeplearning]用户问题优先级分类

2024/01-2025/01

项目负责人

项目描述

根据用户的问题描述，利用基于CNN的transformer模型来做问题优先级分类 (Urgent/High/Medium/Low). 准确率 92%。分类之后，利用问题的优先级在Service Now 网站里提交问题表单。用于用户问题跟踪。

[Java] Spring Cloud Config 集成与配置中心重构

2024/08-2024/12

项目负责人

项目描述

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] Springboot2.x 到Springboot 3.x升级和JDK17 升级

2024/06-2024/12

项目负责人

项目描述

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-Cloud] DevStream 自动化 CI/CD 系统

2023/05-2024/12

框架设计

项目描述

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] Watch Tower

2024/02-2024/06

项目负责人

项目描述

主导开发 Watch Tower 平台，在Red Hat OpenShift 私有云容器平台上对所有微服务的集中可视化管理。

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。

负责系统设计、数据采集、界面开发与后端编排，帮助团队从命令行切换到可视化操作，显著提升运维响应速度与准确性。

[Java-Cloud] 微服务上私有云的迁移

2023/06-2024/05

项目负责人

项目描述

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。

为保障业务连续性，制定“混合部署”方案：老系统在迁移阶段支持双环境运行（Linux + OpenShift），新开发项目则统一部署在私有云中。

[Java] Tibco Queue 到 Kafka & Hessian 到 REST API 迁移改造2023/07-2024/04

项目负责人

项目描述

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 17 环境下因 Jakarta EE 标准兼容性问题而无法使用的技术瓶颈。

[Java-Cloud] OpenShift Pod 自动弹性伸缩系统2023/02-2023/05

项目负责人

项目描述

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。

[Java-Micro] 传统系统微服务化重构2018/02-2019/03

项目负责人

项目描述

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9852112019/06-2022/06

软件工程（人工智能方向）

硕士

延安大学2007/06-2011/06

计算机科学与技术

本科

语言能力

英语（读写精通）；普通话