



Harry Shao

Backend Developer / Full-Stack Developer / AI Agent Developer

中国 / 上海 · shaohaizhou36@gmail.com · github.com/shzharris · timewait.cc/zh/aboutme

做了 10 年企业系统，长期和 ERP、金融、电商这些复杂业务打交道。2025 年开始把主要精力放在 AI 应用落地：给现有系统接入 Agent、RAG 和 MCP，并处理权限、审计、稳定性这些上线后绕不开的问题。

目前关注企业 AI、Agent 平台和业务系统智能化相关机会，工作地点以中国上海、新加坡、日本或远程为主。

技能

主要方向

- Agent / Tool Calling / MCP
- RAG / 文档解析 / PGVector
- LLM API / 本地模型接入
- 权限控制 / 审计 / 可观测性

长期使用

- Python / Go / PHP 后端
- PostgreSQL / Redis / MySQL
- Docker / CI/CD / K8s
- Vue / React 前端协作

工作经历

Black Flame Inc. · AI Agent Developer

2023 — 至今

负责医疗 ERP 智能化：主导订单/生产单 Agent 架构，设计 MCP 工具层、RAG 知识库与权限审计；带领团队完成试点上线，将自然语言查询接入 72 家生产中心日常运营流程。

信飞科技 · Backend Developer

2018 — 2022

主导金融科技 CRM、OMS、催收系统研发。2025 年基于该领域经验独立完成催收平台 MCP 化与 Collections Agent 原型，验证 Agent 在合规催收场景的可行性。

飞牛科技 · Full-Stack Developer

2016 — 2018

负责电商 CMS、OMS、营销自动化系统。2025 年将运营流程抽象为 Marketing Ops Agent + MCP 工具链，形成可演示的智能化改造方案。

AI 应用案例

医疗 ERP 订单智能助手 生产落地

2025

为一家医疗 3D 打印厂商的 ERP 增加订单助手，服务 72 家生产中心。负责 Agent 架构和落地：订单、生产状态与客户档案通过只读 MCP 工具开放，操作手册和 SOP 进入 RAG 知识库。覆盖订单检索、生产进度和异常工单解释，日均调用 200+ 次，平均响应低于 3 秒；相关人工咨询约减少 40%。

AI Agent · MCP · RAG · ERP · Python

金融催收平台 · MCP 与 Agent 改造 独立验证原型

2025

基于 2018—2022 年负责过的催收系统业务模型完成脱敏原型，不连接原公司生产数据。将客户画像、案件状态、拨号策略和合规话术封装为 MCP 工具，支持自然语言查案、分案建议和催收摘要，验证 Agent 在强权限业务中的边界。

MCP · AI Agent · Collections · Redis · Python

电商营销系统 · 运营 Agent 改造 独立验证原型

2025

基于早期电商 CMS / OMS / 营销自动化经验，完成脱敏 Marketing Ops Agent 原型。通过 MCP 串起商品、库存、活动和投放数据，支持文案生成、排期建议、效果问答与异常归因，验证运营流程如何拆成可复用工具。

AI Agent · MCP · CMS · OMS · Marketing

技术写作

- LMCache 实用指南：vLLM 如何复用 KV Cache — 推理服务、缓存与性能
- MinerU 使用指南：文档如何进入 RAG — 文档解析与知识库输入质量

