

苗静思

Agent 开发工程师 · 全栈

+86 175 1209 0401 ✉ miaojsi@outlook.com 🌐 kylinmiao.me 📄 [GitHub](#)

近 4 年腾讯后端开发经验，主导 AI Agent 审核系统 (V1→V5 五版迭代) 与策略全链路平台研发。以 Python/C++/Go 为核心，专注 Multi-Agent 编排 (LangGraph)、RAG/LLM 应用落地与分布式系统设计。兼具 Web 全栈与 iOS 独立开发经验。

开源贡献：AI 全栈开发平台、Agent 记忆框架。

深耕 Agentic Coding，实践从 Vibe Coding 到 Spec-Driven 到 Harness Engineering 的范式演进。

EXPERIENCE

腾讯-企业微信

Jul 2022 – Present

后端工程师 · Agent 开发

主导 3 大核心系统建设：AI Agent 审核 (日均 300+ 单)、策略全生命周期平台、反作弊治理系统 (年化 4000 万)，深度应用 RAG/大数据/分布式系统

SKILLS

技术栈

技术栈以 Python/C++/Go 为核心，深入实践 Multi-Agent 系统设计 (LangGraph/LangChain) 与 LLM 应用落地 (LiteLLM/Langfuse/RAGAS)。熟悉分布式系统与大数据处理 (ClickHouse/Spark/TDW)。具备 RAG 系统全链路经验 (检索策略、Prompt Engineering、评估闭环)，兼具 Web 全栈与 iOS 移动端开发经验。

PROJECTS

风控审核大模型 AI Agent · 负责人

企业微信 Dec 2025 – Present

智能审核系统 (V1→V5 五版本迭代)，支撑企业微信日均 300+ 申诉自动处理

- 主导 V1→V5 架构演进：V1 朴素 RAG → V2 CBR → V3 Prompt Chaining (Token ↓88%，准确率 70%→90%+) → V4 Multi-Query + 策略短路 → V5 自研 **Phase-Gated Constrained ReAct** 三阶段受控 Agent 运行时 (Evidence→Classification→Verdict 流水线 + 10 个 MacroTool + 5 条确定性规则收口)
- 设计双维度 Skill 路由系统 (CaseType + 状态条件)，5 个 YAML 声明式 Skill + 12 个 Advanced Section 按场景动态注入；自研纯 Go BM25 知识引擎 (中文 bigram 分词，530+ 条审核规则 <5ms 检索)；errgroup 并行取证 3 路 MacroTool
- 构建三层质量保障：Self-Review 矛盾检测 + 证据充分度双重验证 (LLM 自评 × 系统硬检) + PolicyArbiter 确定性规则收口；Langfuse 全链路可观测 (三阶段 Span 层级 + Trace Validation YAML spec 自动化验证)；人脸场景 AI 识别率 5-10% → 94.89%
- 设计 Multi-Query 并行检索 + 跨工单历史记忆 (BlockUin 关联)，召回率提升 20%，整体覆盖率达 50-60%，同一用户重复申诉决策一致性显著提升
- 实现分布式高可用审核系统 (DB 原子抢占 + 30s 补偿循环)，支撑 2000+ 并发，日均 300+ 单自动结单，节省 2+ 人力/天

自建应用反作弊治理系统 · 负责人

企业微信 Jan 2023 – Jun 2024

大规模智能检测与自动化治理系统，覆盖 **311 万** 活跃用户

- 主导跨部门项目（安全、产品、法务团队协作），识别违规服务商 **820 家**、企业 **5.2 万** 家，处理亿级实体数据
- 实现年化收益近 **4000 万**，通过多维度算法挖掘（域名 **178 万/月** + IP **38 万/月** + Path 关键词 **30 万/月**）
- 构建 **30+** 模块端到端自动化决策系统（识别→灰度→迁移→反馈闭环），人工干预减少 **80%**
- 协调法务团队建立服务商违规条款，妥善处理百余家企业反馈，实现零投诉运营

策略全生命周期管理平台 · 负责人

企业微信 Jun 2022 – Jun 2025

整合画像中台、图能力、策略质量保障、特征系统的一站式策略管理平台，支撑反垃圾安全治理

- 主导架构设计 **10 亿** 级画像中台（OLAP ClickHouse + OLTP TDW/MySQL），协调数据、产品团队，支持 **200+** 标签 5 分钟复杂分析，刷单欺诈 **↓30%**
- 与基础架构团队协作构建图数据库（**20+** 实体，**70+** 关系），支持 4 种扩散 + 4 种溯源分析，完成 **17 项** 业务入图
- 主导策略质量保障平台（自动挖掘→分级→闭环），挖掘 **228 个** 低质量策略，Review 时长 **5天→2天**，误封率 **2.4%→0%**
- 修复 PKV 分布式序列化缺陷，写入性能 **90k/s→1200w/s**（**130 倍** 提升），解决百亿级数据写入瓶颈

Atoms Plus · AI 全栈开发平台

开源项目 Feb 2026 – Present

AI 全栈开发平台，集成 LLM 智能路由、深度研究引擎与沙箱隔离执行

- 实现 **LLM 智能路由** 系统：LLM 意图分析 → **8** 角色自动匹配（含 Vibe Coding 优先模式），关键词匹配作为降级方案，结合 **15** 个 Microagents 按 Trigger 动态注入系统 Prompt
- 开发 Deep Research 多轮反思引擎：LLM 生成报告结构 → 并行研究（Semaphore 控制 **3** 并发）→ Reflection Loop 识别空白 → 补充搜索，支持 Tavily/DashScope 双搜索引擎
- 设计 CodeAct 执行引擎（Think→Act→Execute→Observe **ReAct** 循环）+ Condenser 上下文压缩 + 双沙箱隔离（本地进程 **<100ms** / Daytona 云端 VM），搭建 FastAPI + WebSocket 实时通信架构

OpenMemory Plus

开源项目 Feb 2026

AI Agent 记忆框架，支持 Multi-Agent 团队协作

- 设计 xMemory 四层架构（Theme→Semantic→Episode→Message），融合文件系统与向量数据库，支持 **4+** IDE Agent
- 实现 Multi-Agent 团队记忆（Team Shared + Agent Private）与 Handoff 协议

BookAnything · Harness Engineering 实践

开源项目 Apr 2026 – Present

将任意源码仓库自动转化为交互式技术书的 AI Harness 系统

- 实践 Harness Engineering 理念 —— 不优化 prompt，而是定义评分标准：构建确定性三维度评估框架驱动 AI 在闭环中自主迭代，将「如何写好 prompt」转化为「如何定义好坏」
- 职责分离架构：图算法做结构规划（社区检测 + 拓扑排序）、AST 做代码理解（**7** 语言零 token）、LLM 只做需要创造力的事；全栈交互式阅读站（知识图谱可视化 + SSE 实时进度 + 命令面板）

磕线 · 负责人

独立项目 Sep 2025 – Present

国内攀岩训练领域头部 iOS 应用，内测阶段 **800+** 用户，月活 **150+**

- 从 0 到 1 完成产品设计、UI/UX、前后端开发、AI 模型训练、DevOps 全流程交付
- 设计 VideoMind AI 智能剪辑流水线，实现高光剪辑与完攀判定，边缘推理成本 **< 0.001 元/视频**，支持 **50+** 并发
- 获法国头部攀岩品牌 YY Vertical 收购邀约

EDUCATION

华南理工大学

Sep 2019 – Jun 2022

硕士 · 计算机科学与技术

中国民航大学

Sep 2013 – Jun 2017

学士 · 电气工程及其自动化

PUBLICATIONS

UAV Visual Navigation System based on Digital Twin — IEEE 18th International
Conference on Mobility, Sensing and Networking (MSN 2022)

2022-12-14

INTERESTS

健身 — 2023 环球广州新秀职业赛 第 4 名 | 攀岩 V3