

中国，上海

王凯睿

AI 应用 / Agent 系统 / 全栈工程师

8 年+ 工程经验，聚焦 Agent 系统设计与 LLM 应用落地，擅长将业务流水线抽象为稳定可控、可自我升级迭代的 Agentic Workflow。

关注 Agent 系统 / 多智能体应用、LLM 应用工程、AI 自动化平台与全栈交付岗位

Email
NeoWang9999@163.com

Phone / WeChat
15321354010

个人网站
[在线简历](#)



8+

年工程经验

3

人 AI 小团队管理

Agent 系统

多 Agent 协作 · 自我升级 · 生产落地

PROFILE

从模型研发到产品落地

我长期在 NLP 模型训练、AI 应用开发、金融文本处理、知识图谱、Agent/RAG 和全栈系统之间工作。擅长将业务流程抽象为可执行的 Agentic Workflow，设计 Agent Loop、工具调用、可检索记忆、状态管理和多步骤任务编排，并用 Python、TypeScript、Docker、数据库和模型服务部署到真实业务场景。

SKILLS

核心能力

Agent 系统

Agent Loop

Tool Calling

Multi-Agent 编排

Memory 管理

分层记忆

RAG 检索增强

Prompt/Context 组装

任务状态管理

工作流编排

评估-反思-重试闭环

Agent Contract

AI / NLP

NLP 模型训练

AI 应用开发

Prompt Engineering

知识图谱

BERT 微调

信息抽取

情感分析

工程与交付

Python

TypeScript

React

FastAPI

PostgreSQL

MySQL

Redis

MongoDB

Neo4j

Hatchet / Celery

Telethon

MinIO

Docker

Linux

GitLab CI

miniqt / ptrade

具备从算法研究、数据标注、NLP 模型训练到 Agent 系统生产落地的完整工程经验。

擅长 Agent 元能力设计：多 Agent 编排、分层记忆、工具调用、评估-反思闭环，以及工作流与状态机治理。

在内容运营自动化、Bug 修复 Agent、游戏 NPC、舆情分析与金融文本等场景有端到端交付经验。

EXPERIENCE

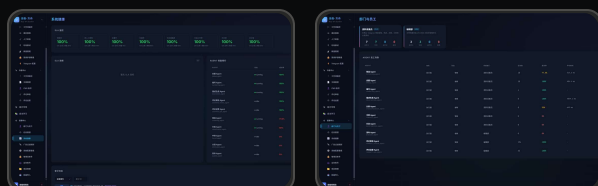
工作经历

2024.10 - 至今

Agent 开发工程师

蝴蝶效应 · 远程

蝴蝶效应是一家专注线上内容运营的公司。我负责 AI Agent 系统的设计与工程落地，将内容运营与研发流程中的重复性工作抽象为可编排的 Agent Workflow。



OASH (方舟) — 运营内容自动化 Agent 调度中枢

Python / FastAPI / PostgreSQL / Redis / Hatchet / Celery / Telethon / MinIO / Docker / GitLab CI

项目概述：公司线上内容运营核心后台。覆盖 Telegram 素材采集、筛选、去重、描述、分类、多频道分发与 CMS 发布，以 Hatchet 事件驱动 Agent 流水线编排任务，配套能力中心、知识服务与 Agent 记忆，实现运营全流程可编排、可观测、可迭代。

我的职责：独立负责 Agent 架构、工作流编排、Contract 治理、记忆/知识闭环与生产部署运维。

- 元认知闭环：AgentContract 约束 I/O 边界；执行成功强化记忆、失败自动生成教训；人工纠正经 knowledge_service 沉淀为规则草稿并审核激活，形成评估-反思-优化迭代。
- 分层经验记忆：情节记忆（FlowLog/执行轨迹）、语义记忆（规则/标准/教训库）、程序记忆（Contract + 状态机 playbook）；执行前 memory_service 按 Agent 上下文召回注入。
- 多 Agent 协作：Hatchet DAG 串联 Screening/Dedup/Description/Classification 等 Agent，Dispatch 扇出与 ParentAgent 子任务分解；能力中心 Kaitian 主-子 Agent ReAct 委派 + Skill Market 工具组合扩展能力边界。

- 自我调试与修复：MaterialStateMachine 幂等转换与 ERROR 回滚；Cron 流水线卡死恢复、分发核验与健康巡检；Hatchet + Celery 双轨任务与 GitLab CI 生产部署。

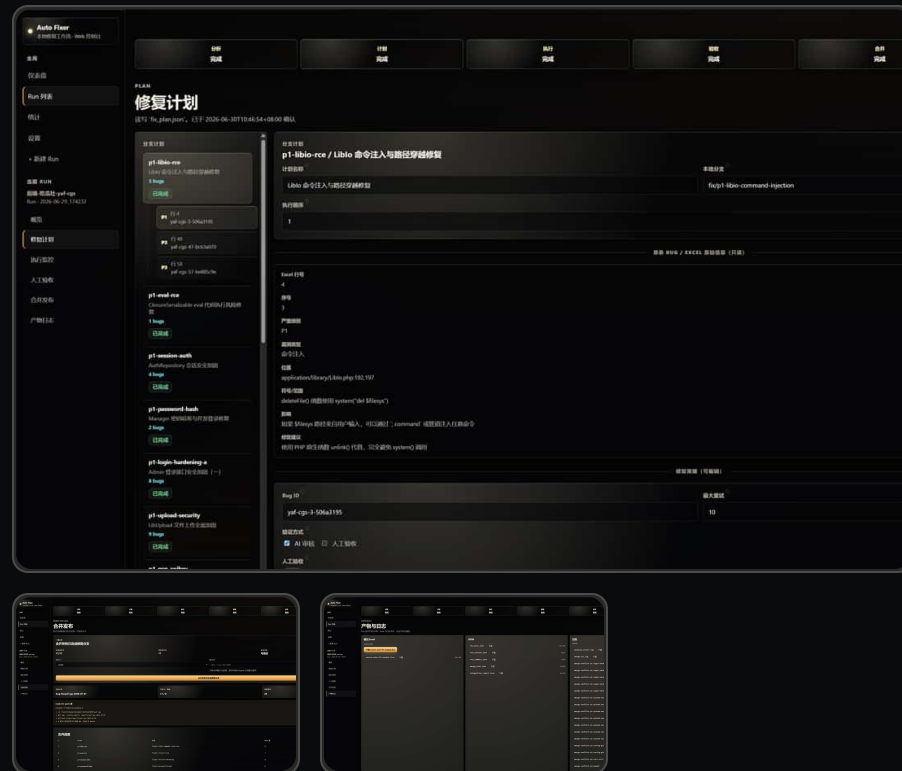
Auto Fixer — Bug 自动发现与修复 Agent 系统

Python / FastAPI / Typer / Pydantic / React / Claude·Codex·Cursor Agent CLI / Git Worktree

项目概述：本地 Bug 发现与修复 Agent 系统。Discover (规则扫描 + AI 复核) 输出审计 Excel/飞书表，Analyze/Fix 读表并行修码，形成发现→修复→审核→经验回流的自进化闭环。

我的职责：负责 Agent 元能力架构：多阶段状态机、统一执行器抽象、分层记忆与「评估-反思-重试」闭环设计。

- 元认知闭环：fix → AI review → retry/resume，将审核结论回注 prompt 诊断失败根因并迭代策略；人工闸门 + Worktree 隔离平衡自进化与生产稳定。
- 分层经验记忆：情节记忆 (run/result.json)、语义记忆 (漏洞模式知识库)、程序记忆 (fix_steps playbook)，在 Discover/Analyze 阶段按需召回。
- 多 Agent 协作：Discover 复核 / 修码 / 审核角色分工，Worktree 并行多分支；统一 Claude·Codex·Cursor 执行器，支持工具链切换与组合。
- Discover 多层发现：Repo Profile + 本地规则 + lint/build + AI 批复核，结构化输出审计表，接入 Excel/飞书修复链路。



2022.05 - 2024.09

AI 部门负责人

竹子网络 · 上海

竹子网络是一家约 300 人规模的游戏研发公司。我负责 3 人 AI 部门，主导游戏中 AI 功能的设计与落地。

游戏 NPC 智能对话与 Agent/RAG 系统

项目概述：面向角色扮演和剧情交互场景，构建基于 LLM + RAG 的 NPC 智能对话系统，将角色人设、世界观设定、历史对话和游戏状态接入可检索记忆与上下文决策流程。

我的职责：负责整体 Agent 架构设计与工程实现，设计 NPC Agent Loop、Memory 检索、上下文组装、模型服务接入和游戏侧调用流程。

- 将玩家输入、角色设定、世界观知识、历史对话、游戏状态和行为选择纳入统一 Agent Loop。
- 设计可检索 Memory 机制，对角色设定、剧情知识和历史对话进行召回、裁剪与上下文组装。
- 将对话管理、知识检索、状态判断和行为选择编排为多步骤 Agent Workflow，提升 NPC 回复一致性、角色语气稳定性与玩家沉浸感。

动作识别与动作风格迁移

项目概述：基于用户上传的视频，进行骨骼关键点检测，对角色动作进行识别，并将动作迁移到游戏角色上。

我的职责：负责算法方案验证、关键点检测流程接入、动作特征处理和效果调试，提供稳定的 API 服务。

- 将算法能力封装为可被游戏功能调用的工程模块，便于后续功能集成。

2020.07 - 2021.07

NLP 全栈工程师

金证股份 · 北京

金证股份为证券、基金等金融机构提供核心系统与数据服务。我负责 NLP 算法研发，服务香港交易所与上海证券相关客户项目。

香港交易所合同信息抽取系统

项目概述：为香港交易所研发合同信息自动化抽取系统，支持对合同文档中指定文本、段落或章节进行精准提取，并结合规则引擎与机器学习模型构建可配置的信息抽取平台。

我的职责：负责信息抽取平台后端架构设计，与产品经理协作设计用户操作逻辑和前端页面交互，并制定开发排期与功能边界，完成深度学习的模型训练，保障项目按时交付。

- 将合同文档抽取需求拆解为可配置的规则、模型和交互流程，提升平台可维护性。

上交所金融知识图谱

项目概述：面向证券信息检索场景，构建以公司为中心的金融知识图谱，将通用预训练模型本地化，并针对金融领域数据进行微调优化。

我的职责：负责知识图谱模型开发与图数据库建设，使用 ArangoDB 与 Neo4j 构建图存储，开发图查询算法，并协调后端数据库 API 协议设计与前端联调。

- 围绕金融实体、关系和检索路径建设图谱能力，提升证券信息检索效率。

2017.07 - 2020.06

NLP 全栈工程师

阿博茨科技 · 北京

阿博茨科技专注金融文本智能处理，为上市公司公告与财务报告提供结构化数据服务。

上市公司公告信息抽取

项目概述：利用机器学习算法对上市公司年报进行文档理解与关键信息抽取，支持股权转让、董监高人员变动、股价变动等多类事件的自动识别与结构化输出。

我的职责：负责年报语料组织与标注，制定标注指南和规范文档，完成训练数据生产与质检；基于 Google 开源 BERT 模型进行分类任务预训练与微调，通过训练文本分类和序列标注模型，解决真实项目中的信息抽取问题。

- 从数据标注规范、训练数据质量、模型微调到服务 API 搭建，形成可交付的金融公告信息结构化流程。

金融数据结构化 SaaS 平台

项目概述：面向金融机构的信息结构化平台，支持用户上传 PDF、Word 文档和图片，自助选择预设数据结构与机器学习模型进行进一步训练和测试。

我的职责：开发 PDF 转 Word 模块和模型训练调用 API，推动移动端内部结构预测模型的客户部署，并在 Linux 集群上使用 Docker 完成算法服务的容器化部署与运维。

- 打通文档转换、模型调用、客户部署和服务运维流程，提升算法服务产品化能力。

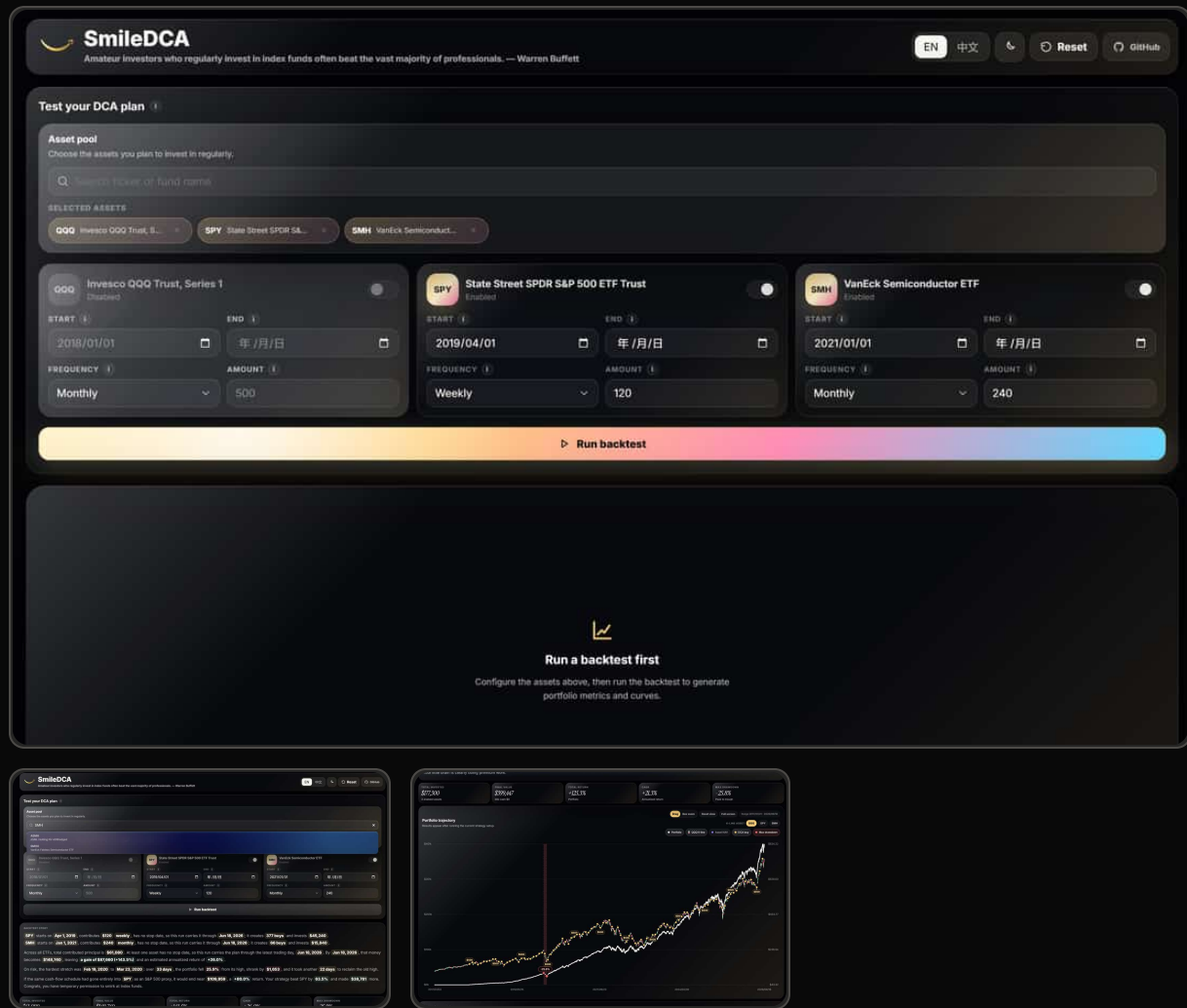
SELECTED WORK

个人项目

我持续开发面向个人投资者和金融数据场景的软件项目，覆盖开源工具、量化交易、财经数据采集、AI 舆情分析和 Agent/RAG 报告生成，并重点打磨任务编排、状态管理和可检索记忆能力。

OPEN SOURCE

个人开源项目



SmileDCA

github.com/NeoWang9999/SmileDCA

Vanilla HTML / CSS / JavaScript / Canvas 2D / IndexedDB / localStorage / GitHub Actions / Cloudflare Pages / Cloudflare Workers / R2

面向 ETF 投资者的轻量级浏览器端定投回测模拟器，支持构建多 ETF 定投计划、运行历史回测，并在一个交互式仪表盘中查看结果。

- 支持按 ETF 独立配置起止日期、定投频率、每次投入金额和启用状态，并生成组合级回测结果。
- 在浏览器端完成回测计算，展示组合净值曲线、单 ETF K 线/NAV 叠加、定投买点标记、最大回撤区间和 SPY 基准对比。
- 内置中英文切换、明暗主题、图表拖拽缩放、全屏查看和策略状态保存，提升个人投资者的交互分析体验。
- 使用静态 JSON 管理 ETF 元数据与行情数据，通过 GitHub Actions 定时更新数据并部署到 Cloudflare Pages；Cloudflare Functions / R2 支持缺失价格文件的按需代理与缓存。

PRIVATE WORK

个人非开源项目

量化交易策略研究平台

Python / TypeScript / React / Lightweight Charts / miniqmt / MT5 API

基于 TradingView 开源 Lightweight Charts 搭建，支持自研指标可视化、策略回测、因子分析和实盘交易切换。

- 绕过 TradingView 自定义指标限制，在本地实现 Python 指标展示、回测与因子分析。
- 策略逻辑可在本地开发后无缝切换到回测或实盘模式。
- 通过 miniqmt 实现股票与可转债 tick 级回测和实盘交易，并具备 MT5 Python API 框架。



AI 舆情分析系统

Python / FastAPI / MongoDB / Playwright / OpenAI / React / Ant Design / APScheduler

持续爬取多个财经新闻源，经 AI 分析汇总呈现，帮助用户快速把握市场情绪与行情信息。

- 支持财联社、知识星球等新闻源的定时与手动爬取，包含 Cookie 登录管理、失效检测、增量和全量模式。
- 支持按公司名称或股票代码检索舆情，并通过 MongoDB 存储大规模新闻数据。
- 设计多步骤舆情分析 Agent Workflow，自动编排爬虫取数、数据清洗、实体检索、情感分析、主题归纳和报告生成。
- 设计任务状态与中间结果管理机制，支持定时任务、手动触发、增量爬取、失败重试和分析报告流式展示。



EDUCATION

教育背景

安徽农业大学

管理学学士 · 农林经济管理专业 · 2017 届毕业