

张恒玮

AI Agent 应用开发工程师 / 大模型应用开发

北邮·新一代电子信息信息技术硕士·2027 届·期望 base 北京 | 17638342407 | heng1537@foxmail.com | github.com/galanime
Chart Agent 已交付内部团队日常使用（单图 30 分钟 → 2 分钟） · 实习 3 个月独立交付 5 个软件产品 · 6 个公开项目可跑通演示

教育背景

北京邮电大学 | 新一代电子信息信息技术 硕士（2027 届） | 学业排名前 10%，校级一等奖学金 ×2、优秀研究生
山东科技大学 | 通信工程 本科 | GPA 3.58/4.0，专业前 6% | MCM/ICM H 奖、市场调查与分析大赛国家级二等奖
成果与语言 | 软件著作权 1 项 · EI 会议论文 2 篇（5G 应用与通信工程方向） | CET-6 通过 · 日语 N2

专业技能

编程语言：Python（主力，实习与全部后端项目）、Go（CareerPilot Runtime，含官方 anthropic-sdk-go 集成）、TypeScript / React（项目前端）、SQL

大模型应用：大模型 API 集成（Anthropic 官方 SDK / OpenAI 兼容 / DeepSeek / Ollama 本地部署）、Provider 抽象与多后端切换、Tool Calling 注册与 JSON Schema 调度、Multi-Agent 阶段化流水线、语义匹配与置信度拒答、幻觉控制（证据驱动 + 输出校验 + 规则回退）

Agent 架构：阶段化工作流编排与 Run / Step / Artifact 落库、人工审批门禁、风控拦截层、审计日志与 Trace 回放、效果评估面板与指标定义；n8n 可视化编排（已用于真实内容生产链路）；了解 LangChain / LangGraph 的定位，项目中出于可控性与可审计性考虑使用自研 Runtime

后端 & 部署：FastAPI、Pydantic v2、SQLite / PostgreSQL / Redis、Celery 异步任务、Playwright 自动化、Docker Compose 多服务编排；Pytest（Python 侧 130+ 用例）与 Go testing

AI 编程工具：Claude Code、Codex（日常主力开发方式）；全部项目由本人设计架构与技术选型，借助 AI 工具完成实现与迭代

Agent 项目经历

CareerPilot Agent — Go 实现的 Agent Runtime（多 Provider 可切换）

Go 1.24 · Anthropic 官方 Go SDK · React 19 + Vite · SQLite（纯 Go 实现，无 CGO） | github.com/galanime/Careerpilot-agent

- ▶ 从零设计 Agent Runtime：Tool Registry 以 JSON Schema 注册 **8 个工具**，编排为八阶段执行链路（解析 JD → 证据检索 → 匹配打分 → 材料生成 → 产出评估 → 档案打包），Run / Step / Artifact 三表落 SQLite 支持全链路 Trace 回放；仅用 Go 标准库 HTTP + SSE，编译为单文件分发
- ▶ 设计 Provider 抽象层，**fake / Anthropic 官方 SDK / OpenAI 兼容 / Ollama 本地部署四种后端可切换**，默认 fake 保证零 Key 零成本可复现；生成环节强制证据驱动，以关键词覆盖率阈值 + 无证据主张检测做输出校验，模型调用失败自动降级规则模板保证链路不中断

闲鱼 AI 智能客服系统 — 大模型驱动的三层 Agent 架构

Python · FastAPI · SQLite · DeepSeek API · uiautomator2 | github.com/galanime/xianyu-ai-customer-service-agent

- ▶ 针对商家客服「对话量大、AI 易做出错误承诺」的真实风险，设计三层架构：LLM 理解层（订单 / 价格 / 押金 / 自提 / 配送 5 类意图识别 + 回复润色）→ RiskGuardAgent 风控层（判定是否需人工接管，并阻断确认订单 / 改价 / 写入知识库 / 未审发送 **4 类高危动作**）→ ADB 执行层（仅填草稿，发送动作需显式开关 + 请求级确认）
- ▶ 业务规则以可替换知识库（business / rules / reply_style YAML + FAQ JSONL）与 Agent 逻辑解耦，同架构可迁移到其他垂类；审计日志完整记录决策 Route Trace 支持会话回放与合规导出，17 项单测覆盖风控判定

Chart Agent — 端侧 AI 出图 Agent（企业实习期间 0 → 1 落地）

Python · SQLite · pdfplumber · Matplotlib | 独立设计与开发，已打包交付内部使用

- ▶ 面向芯片 Datasheet 验证场景设计端侧 Agent：PDF 图题提取 → CSV 列语义匹配 → 模板参数映射 → SVG 自动出图；设置置信度阈值，低置信度直接拒绝生成而非猜测，单张图表交付从人工 30 分钟压缩到 2 分钟
- ▶ SQLite 分层记忆系统（工作 / 情景 / 语义 / 程序记忆），新经验需人工审批后才参与后续匹配；输出 manifest.json + match_report.json 审计清单，证明产物数据仅来自输入 CSV；完全端侧运行、数据不出内网，满足芯片行业数据安全红线

LifeHelper — 全栈 AI Agent 工作台（含效果评估体系）

Python · FastAPI · React + Ant Design · Redis · PostgreSQL · Celery · Docker Compose | github.com/galanime/LifeHelper

- ▶ 实现意图分析 → 任务拆解 → 工具调用 → 结果聚合完整 Agent 链路，8 个内置工具 + 重试机制 + 会话快照，异常场景自动降级；Mock fallback 与 OpenAI-compatible provider 双通道，零成本验证与演示
- ▶ 搭建评估面板 + 人工评分入口 + Trace 回放 + 执行图可视化，定义任务完成率 / 重试率 / 端到端时延作为核心体验指标；**112 项 Pytest 用例**全部通过，Docker Compose 一键拉起 API / Worker / Beat / Redis / PostgreSQL 五服务

实习与工程经历

Python 应用开发 — 苏州领慧芯科技（高性能模拟芯片设计） | 2026.05 - 2026.07

- ▶ 3 个月内独立交付 5 个软件产品：示波器桌面控制应用（10+ 版本迭代 / 21 项自动化回归测试 / Windows x64 打包）、Chart Generator 三产品架构（Web → 桌面端 → Chart Agent，共享引擎独立打包）、ADC 采集 GUI（Mock / CSV / FPGA 三后端可切换）
- ▶ 治理数据完整性风险：VISA I/O 加锁串行化解决导出 / 断开 / 关窗的并发资源竞争，SCPI 二进制块长度校验拒收截断数据，.part 临时文件 + 原子替换确保「文件要么完整、要么不存在」

技术开发兼职（已上线产品 + AI 工作流落地） — 北京灰虎研题教育 | 每年 04 - 08 月

- ▶ 考研题库微信小程序开发并已上线运行，覆盖 **5 所院校约 5000 名** 当届学员，打通学员 / 教师 / 管理员三角色完整流程；搭建小红书渠道 AI 客服自动承接咨询，配合内容运营 **半年全平台涨粉 2 万** 持续导流
- ▶ 用 n8n 编排「选题评分 → 文案 / 视觉 Prompt → 格式校验 → 人工审核 → 失败兜底」内容生产工作流：人工审核为强制卡点，任一节点失败自动回退至上一稳定产物、不中断整条流水线，**单篇制作** 由约 **4 小时** 缩短至 **15 分钟**、**周更** 由 **1 篇** 提升至 **7 篇**