



核心优势

业务到产品: 医疗与 HR 场景实践, 参与 VedAI 上线; 独立推进 Clinical、UniHR 与 EVO, 将需求拆解为流程、验收标准和可运行原型。

评测驱动迭代: 围绕真实执行链路设计分层评测, 用固定样本、原始输出和节点证据定位失败, 区分执行完成、质量通过与待复核。

教育背景

康奈尔大学 医疗信息学与人工智能理学硕士

2024.09–2025.09

西交利物浦大学 / 利物浦大学 生物信息学理学学士

2020.09–2024.07

项目经历

Agent EVO • Agent 评测与失败诊断平台

2026.10–至今

个人项目 | 产品设计、AI Coding 实现 | Clinical 案例贯穿评测链路

- 问题与方案:** 针对“有最终答案却难判断哪里出错”, 搭建测试集→真实业务入口→节点追踪→分层评测→疑似故障诊断→回归链路; 支持测试集上传、批次管理、逐轮回放与原始输出查看。
- 评测设计:** 按意图、检索、来源、产物、上下文与安全拆解验收; 结合规则检查、模型辅助评分和人工复核。关键安全项独立判断, 缺少 trace 或金标时保留证据不足 / 待判断, 避免平均分掩盖风险。
- 诊断与验证:** 将失败断言关联到节点及原始证据, 区分疑似原因与未排除解释; 选中重跑生成新记录, 固定用例、知识与版本快照, 保留历史用于对照。
- 工程结果:** 100 条合成用例 (UniHR / Clinical 各 50) 完成真实业务执行, 取得 **541 次模型调用证据**; 32 条涉及失败断言, 40 项关键安全待判断。明确执行完成≠质量通过, 尚无医生临床有效性验证。

Clinical Agent • 精神科临床辅助工作台

2025.07–至今

基于 VedAI 的个人持续迭代 | 产品设计、原型实现 | dominicspace.com/#clinical

- 产品判断:** 针对多轮病例信息遗漏、医学证据利用不足和过度确定回答, 统一病例与自然语言入口, 复用已确认事实; 以 RAG 和专业 Skills 支撑任务路由及指南调用。
- 安全机制:** 设计 Context、Evidence、Answer Integrity 三阶段核查, 分别验证病例事实、证据支持及回答引用; 证据不足时补检索或收缩回答, 高风险结论保留医生复核。
- 失败到改进:** 针对隐含追问漏识别与“来源正确、药品错误”, 改进语义路由和药品实体重排; 同组 **100 条合成路由样本 38%→98%**, 同组 **20 条固定检索查询 Mean Precision@5 57%→93%**。两组均为离线工程对照, 不代表临床效果。

UniHR AgentOS • HR 任务与文件交付工作台

2026.05–至今

个人项目 | 产品规划、功能设计与实现 | dominicspace.com/#unihr

- 任务与验收:** 将政策查询、公文与表格处理组织为可查看、可修改的任务计划, 按依赖编排工具执行; 依据政策来源、文档字段及表格公式验收 Word / Excel 产物, 关键操作保留人工确认。
- 迭代基础:** 完成自然语言与附件输入到文件交付的原型链路, 构建 **50 项固定测试用例**, 覆盖路由、协作、工具异常和确认节点; 通过任务回放定位失败并调整规则。

AI 相关工作与实习实践

浙江大学“一带一路”国际医学院

2025.12–至今

行政管理岗 | 内部 AI 提效实践

- 通过经办人访谈梳理人事、行政重复操作与跨文件协作, 按业务价值、规则稳定性和风险筛选 AI 场景; 对齐文件模板及确认节点, 收集异常反馈, 培训同事使用提示词、Workspace 与 Skills。

卡罗莱纳精神病研究所 • Carolinas Psychiatric Institute

2025.01–2025.07

AI 临床辅助产品实习 | VedAI

- 参与医生访谈与病例复盘, 整理 **200+ 临床场景与 Prompt 交互**; 参与需求与 Agent 流程设计, 借助 AI Coding 搭建 Demo, 将医生反馈转为输出标准及交互需求, 参与推动 VedAI 上线。

专业技能

产品与评测: 需求访谈、场景拆解、PRD、原型、验收标准、固定测试集、版本对照、失败诊断与人工复核。

技术与语言: Python、AI Coding / Codex、LangGraph、RAG、Routing、Tool Use、Skills、MCP、HITL; 雅思 7.0, 英文技术阅读与沟通。