

教育经历

南方理工大学 省部共建高校 卓越工程师教育培养计划 新工科研究与实践项目 计算机科学与技术 / 工学学士	2021.09 - 2025.06
学位授予方：南方理工大学	
- 主修机器学习、分布式系统、编译原理与数据库系统；连续三年参与智能软件工程实验室 Agent Systems 方向。 - 学校 Title 仅采用学校官网与教育主管部门公开表述，不把校企合作、交换访问或实验室合作写成学位归属。	

实习 / 工作经历

星桥科技 AI4SE / Agent Scaling & Infrastructure SFT / RL / Agent Harness / Context Engineering / SWE-Bench	2025.07 - 至今
AgentScaling for SFT & RL - 轨迹合成模块 Owner 0→1 多 Scaffold 数据管线 训练-评测闭环	
背景：	单一模板生成的 Coding Agent 轨迹同质化明显，难以覆盖规划、工具误用、失败恢复与长程续跑等能力维度。
指标与效果：	- 目标是在同一任务上生成多 Scaffold、多模型、多采样温度的异构轨迹，并保留可复现的环境、补丁和评测工件。 - 累计合成 180k 条 Micro Trajectory 与 42k 条 Long-horizon Trajectory，覆盖 11 类失败模式和 8 类工具链路。 - 固定 1,200 题回归集上，训练后模型的任务通过率从 23.8% 提升至 31.6%，平均无效工具调用从 4.7 次降至 3.1 次。
我的职责：	- 通过容器快照复用、仓库依赖缓存与分层超时，将单条有效轨迹平均生产成本降低 27.4%。 - 作为轨迹合成模块 Owner，从 0→1 设计 CLI -> Scaffold Registry -> Relay Runner -> Sandbox -> Verifier -> Exporter 六阶段管线，定义统一 task schema、artifact contract 与 retry policy。 - 设计 Multi-Scaffold 采样：Plan-and-Execute、ReAct、Manager-Worker、Critic-Refine 与 Test-driven 五类 Agent 拓扑共享环境层，但隔离 system prompt、memory 与 tool budget。 - 实现 Docker/Firecracker 双运行时、仓库还原、依赖预热、patch capture、checkpoint resume 与失败回收，覆盖 clone/build/test/edit/export 全生命周期。 - 建立 Trajectory Quality Gate：按 Outcome、Effective Action Ratio、Patch Validity、Token Efficiency 与 Failure Onset 组合过滤，向 SFT/RL 训练侧输出分层数据。
技术关键词：	SFT / GRPO / Multi-Scaffold / SWE-Bench / Docker / Firecracker / Checkpoint / Patch Validity / Effective Action Ratio / Trajectory Synthesis / Quality Gate
Agent Tracer & Rubric Evaluation - 评测诊断模块 Owner Evidence-to-Action Gap Wash Pipeline	
背景：	传统 pass/fail 指标无法解释 Agent 在何处偏离有效路径，也无法区分规划错误、工具错误、环境错误和补丁质量问题。
指标与效果：	- 在 5 个模型、4 套 Scaffold、9,600 条轨迹上完成 Rubric Eval，人工复核一致率 86.4%，高风险失败样本召回率 91.2%。 - 将失败定位从人工逐轮阅读缩短为平均 2.8 分钟/条，支持训练前洗数、版本回归和模型间行为对齐分析。
我的职责：	- 从 0→1 定义 Trace Schema，统一 USER/ASSISTANT/TOOL 消息、observation、artifact、patch、test result 与 checkpoint，并构建跨模型 Adapter。 - 实现 LLM-as-Judge + Rule Check 双路评测，将 Planning、Tool Selection、Error Recovery、Patch Quality 与 Evidence Usage 拆成 29 个原子 Rubric。 - 提出 Failure Onset 定位：从结果失败向前回溯首个不可逆错误，并以 Evidence-to-Action Gap 区分“已观察但未行动”和“根本未观察”。 - 搭建 Wash -> Diagnose -> Export 流水线，支持多维 Top-K、去重、hard-negative 筛选、Rubric 证据片段导出与训练数据回写。
技术关键词：	LLM-as-Judge / Rubric Eval / Failure Onset / Evidence-to-Action Gap / Trace Schema / Hard Negative / Wash Pipeline / Inter-Annotator Agreement / Error Taxonomy
Long-horizon Context Continuation - Context Engine 共建者 Micro Compaction Checkpoint Resume	
背景：	长程 Coding Agent 在 180k token 后出现上下文拥塞、重复探索 and 工具结果遗失，需要在不中断任务的前提下压缩并续跑。
指标与效果：	在 240 个超长程任务上，将可持续运行窗口从约 180k token 扩展到 430k token，中断恢复成功率达到 88.7%。
我的职责：	- 共建 Write -> Select -> Compress -> Isolate -> Resume 上下文链路：区分原始消息、工具工件、摘要记忆、任务状态和失败报告五类载荷。 - 实现 Micro Compaction 与 Full Compaction 双阈值策略，在保留 active plan、open todos、latest patch 与 test evidence 的同时裁剪冗余观察。 - 负责 Checkpoint Resume 的状态一致性验证、悬空 ToolCall 修复与跨模型 handoff 回放，不将该共建职责表述为整个平台 Owner。
技术关键词：	Context Engineering / Micro Compaction / Full Compaction / Checkpoint Resume / Memory / Tool Artifact / Handoff / State Consistency

云岚视频 直播与电商 / Agent 产品算法 Agentic Workflow / GenUI / EC Analysis / Function Calling	2025.01 - 2025.06
直播间智能搭建平台 - Workflow 策略模块 Owner 意图到 UI 的端到端生成	
背景：	直播运营配置依赖人工拖拽组件和填写规则，页面搭建耗时长，跨活动复用成本高。
指标与效果：	- 上线后覆盖约 46% 的新建与改版页面，单页首版搭建时间从 2-4 小时降低至 15-30 分钟。 - 支持 14 个基础组件与 38 类动作绑定，生成结果通过 Schema、DSL 与 Runtime 三层门禁后进入预览。

我的职责：

- 作为 Workflow 策略模块 Owner，从 0→1 设计 Prompt Router -> Intent Planner -> Task Graph -> XML Patch -> Preview Renderer 链路。
- 构建 UI/Flow/Bind 三类任务规划器，将自然语言意图、页面 UIDL 与组件能力表映射为结构化 task list。
- 实现 Promise.allSettled 并行调度、XML Patch 增量更新、JSX/XML/UIDL 多级兜底与失败回滚。
- 建立 Component Schema 与 Action Binding 训练语料，为后续 Semi/Auto Design 组件库奠定可检索能力描述。

技术关键词： GenUI / UIDL / Task Graph / XML Patch / Function Calling / Schema Registry / Action Binding / Promise.allSettled / Fallback

EC-Monica 电商经营分析 Agent - 分析编排模块 Owner | 指标归因 | 报告生成

背景：

- 面向产品、运营和研发的电商分析需求，需要同时处理指标查询、业务归因与需求文档解析，并生成可复用报告。

指标与效果：

- 月均被调用 600+ 次，报告类任务占比约 55%，需求归因解析约 45%；单次分析报告生成从 1-2 小时缩短至 5-15 分钟。

我的职责：

- 从 0→1 建立 Intent -> Query Plan -> Tool Registry -> Data Fetch -> Analysis -> Report 六阶段 Agentic Workflow，支持意图歧义与数据缺失回退。
- 设计 FunctionSchema 动态注册与 Loop Executor，注入数据表口径、业务词典、历史报告 Few-shot 和分析模版。
- 实现超时重试、SQL/接口错误分类、结果完整性检查与无数据 Fallback，避免模型在缺少证据时生成确定性归因。

技术关键词： Intent Recognition / Tool Registry / FunctionSchema / Query Planning / Business Dictionary / Few-shot / Report Generation / Data Fallback

极昼智能 | 多模态内容平台 / AIGC | MultiModal / CLIP / Diffusion / ReAct

2024.06 - 2024.12

Thumbnail AI - 个人项目 | 0→1 多平台封面生成 SaaS

背景：

- 面向 YouTube、Instagram、TikTok 与小红书创作者，将标题、素材与风格词转为可编辑的多尺寸封面。

指标与效果：

- 累计注册用户 2,080，付费用户 214，月活约 540，累计生成封面 15,600 张；访问到付费转化率 10.3%。
- 用户从上传素材到生成首版平均耗时 92 秒，支持 6 个平台、18 种尺寸和 24 套风格模版。

我的职责：

- 独立完成 Next.js 前端、支付订阅、对象存储、任务队列与图像生成服务，打通标题/素材 -> Prompt Planner -> Layout -> Diffusion -> Preview 链路。
- 使用 CLIP 提取参考图视觉特征，结合品牌色、人物主体、文字安全区与平台比例生成结构化 Prompt。
- 实现免费额度、订阅、订单、重试、内容审核和失败退款闭环，并对转化漏斗做事件埋点。

技术关键词： Next.js / Stripe / Object Storage / Task Queue / CLIP / Diffusion / Prompt Planner / Layout Engine / Funnel Analytics

Krene-Art 角色设定 Agent - 搜索-理解-生成闭环 | 参考图一致性 | 角色卡片

背景：

- 面向游戏与二维内容创作者，从角色文字设定和参考图出发生成风格一致的角色卡片与多视图草图。

指标与效果：

- 累计用户 1,260，付费用户 168，生成角色卡 4,300+ 张；单角色从参考检索到首版平均 7.2 分钟。

我的职责：

- 设计 Search ReAct：拆解造型、配色、气质与时代关键词，检索参考图后进行清晰度、构图与视觉维度筛选。
- 以 CLIP + Vision Encoder 提取多维视觉特征，与用户设定拼接为结构化 Prompt，再调用 Diffusion 生成多候选角色图。
- 构建 Critic Loop，按轮廓、配色、服装、年龄感与参考图一致性打分，低分样本自动调整 Prompt 并重采样。

技术关键词： Search ReAct / CLIP / Vision Encoder / Structured Prompt / Diffusion / Critic Loop / Reference Consistency / Multi-view

开源贡献

deer-harness Contributor Agent Harness、前端调试台与文档维护；项目累计 7.6k Stars • 提交 9 个 PR，其中 6 个合入：新增 tool-call 调试面板、checkpoint 恢复提示、两组回归测试与部署文档；不将 Contributor 描述为核心作者。 链接：https://github.com/example-agent/deer-harness
eval-trace-core Maintainer Trace Schema、Rubric Runner 与版本发布 • 仓库维护者列表与 CODEOWNERS 均记录 Maintainer 身份；负责 0.4-0.7 四个版本的 Schema 迁移、发布说明和兼容性评审。 链接：https://github.com/example-agent/eval-trace-core
fast-transport Contributor Python Binding 与序列化 Benchmark • 合入 3 个 PR，修复 Python Binding 的 buffer lifetime 问题并补充跨语言 benchmark；角色限定为 Contributor。 链接：https://github.com/example-agent/fast-transport

技术项目与沉淀

Agent Systems Notes 技术内容与课程共建 Agent Harness、Context Engineering、SFT/RL 与 Evaluation 系列教程 • 整理 42 篇技术笔记、18 个可运行 Demo 和 6 次公开分享，全部示例绑定对应仓库提交、论文或官方文档。 链接：https://example.com/agent-systems-notes
--

Open Benchmark Service | 项目发起者 | 0 → 1

SWE、Search 与 Tool-use 三类 Benchmark 的统一容器化评测服务

- 实现 CreateSession -> Restore Repo -> Run Agent -> Diff -> Test -> Export 标准协议，支持 5 个模型 Provider 与 4 类 Agent Scaffold。

链接：https://github.com/example-agent/open-benchmark-service

Agent Visual Trace Studio | 项目 Owner | 可观测与诊断

把多轮 Agent 轨迹转为可筛选、可回放、可归因的前端诊断工作台

- 实现 Session Timeline、Tool-call DAG、Patch Diff、Token/Latency Flame Graph 与 Failure Onset 高亮，支持 20 万级消息的分片加载。
- 打通 OpenTelemetry -> ClickHouse -> Trace Query DSL -> React Virtual List，支持按模型、Scaffold、Rubric、仓库与失败类型交叉过滤。

链接：https://github.com/example-agent/agent-visual-trace

SWE-Bench Adapter Pack | 核心维护者

统一仓库准备、测试发现、Patch 应用与结果导出的 Benchmark Adapter

- 覆盖 Python、Java、TypeScript 与 Go 四类仓库，抽象 RepoSpec、TestSpec、PatchSpec 与 ArtifactSpec，减少不同 Benchmark 的胶水重复。
- 加入 hermetic build、依赖镜像指纹、flaky test 重跑和污染检测，使同一 patch 的跨节点复跑一致率提升至 97.6%。

链接：https://github.com/example-agent/swe-bench-adapter-pack

Context Compression Playground | 实验平台发起者 | 0 → 1

面向长程 Agent 的 Selective Memory、Compaction 与 Resume 策略对比平台

- 内置 Sliding Window、Summary Memory、Entity Memory、Task-state Snapshot 与 Retrieval Memory 五类基线，统一衡量 Recall、Drift 与续跑成功率。
- 提供 token budget sweep、context ablation、needle replay 和跨模型 handoff，对摘要丢失、状态冲突与悬空工具调用进行可视化对比。

链接：https://github.com/example-agent/context-compression-playground

Agent Engineering 公开分享 | 讲师与内容作者

从 Agent Demo 到可训练、可评测、可续跑系统的工程实践

- 完成《Harness 不是 Prompt》《从 Pass/Fail 到 Failure Onset》《Long-horizon Context Continuation》三场公开分享与配套 Demo。
- 课程代码包含 12 个渐进实验：Tool Registry、Trace Schema、Rubric Judge、Wash Pipeline、Checkpoint 与 Compaction。

链接：https://example.com/agent-engineering-talks

论文、专利与技术沉淀 | 共同作者

Agent 轨迹诊断、长上下文压缩和工具调用可靠性

- 共同完成 1 篇 workshop paper、2 项发明专利申请与 4 篇内部技术 RFC；个人贡献限定为实验平台、评测方法和工程实现。
- 不使用“首创”“世界第一”等缺少比较全集的最高级，公开材料按正式作者顺序和已公开状态呈现。

链接：https://example.com/publications

奖项与技能

全国大学生软件创新赛 / 全国一等奖 2024.08	Python / TypeScript / PyTorch / Ray / vLLM / SGLang / Docker / Firecracker / OpenTelemetry / PostgreSQL / Redis / Next.js
Open Source Summer / 优秀结项 2024.10	
校级优秀毕业设计 2025.06	