

吴汉卿

具备 AI Agent 全流程实践，掌握工具调用、长上下文治理、记忆机制等技术

TEL: (+86)13020631814 | Email: 654601458@qq.com



教育背景

杭州电子科技大学	硕士	计算机技术	2024.9 – 2027.6
• 杭州电子科技大学学业奖学金 (2 次)；研究与实践方向聚焦 AI Agent、RAG 与大模型应用工程。			
青岛大学	本科	计算机科学与技术	2020.9 – 2024.6
• 荣誉奖项/相关课程：青岛大学学业奖学金；人工智能技术 (96)、计算机视觉 (95)；CET-6 (547 分)。			

项目经历

MiniCodeAgent	本地代码 Agent Harness	2026.3 - 2026.6
---------------	--------------------	-----------------

技术：Agent Harness、Tool Calling、Context Management、Layered Memory、Checkpoint / Resume

项目背景：面向代码仓库长链路任务，解决多轮开发中 prompt 膨胀、重复读文件、状态丢失、工具副作用不可控和结果难复盘的问题，目标是让本地代码 Agent 的执行过程可复盘、可恢复、可评测。

- **【Agent 架构设计】统一执行链路**：负责本地代码 Agent 的整体设计与开发，统一模型接入、工具执行、会话状态、checkpoint 恢复与运行工件落盘，支持 **2 类模型后端、7 类工具和 3 类运行工件**。
- **【长上下文治理】压缩 prompt 成本**：设计分层上下文管理与预算裁剪机制，在 **12 组长上下文配置** 中将平均 prompt 长度从 **7082** 压到 **5664**，平均压缩率 **16.19%**，最高压缩率 **33.28%**。
- **【结构化记忆】减少重复读文件**：将任务摘要、文件摘要、过程笔记和相关记忆召回分层管理；在 **12 个记忆依赖任务** 中，follow-up 阶段重复读文件次数从 **60 次** 降到 **0 次**。
- **【任务恢复机制】保证中断可继续**：设计 checkpoint / resume 机制，使 Agent 在上下文超预算、中断恢复和 workspace 漂移场景下恢复任务状态；覆盖 **10 个恢复场景**，workspace 漂移识别率 **100%**。

智能旅行规划 Agent 系统	RAG / 工具增强规划	2026.1 - 2026.3
-----------------	--------------	-----------------

技术：FastAPI、HelloAgents、Vue3、TypeScript、Pydantic、高德地图 API、SFT / DPO

项目背景：项目面向真实旅行规划，构建智能旅行规划助手以生成可校验的多日行程方案，重点解决大模型直接生成行程时的幻觉、预算不可控和结果不可校验问题。

- **【结构化规划链路】约束模型输出**：构建 TripRequest -> PlannerContext -> TripPlan JSON 流程，将景点、酒店、餐饮、天气、预算约束和价格 hint 统一注入上下文，降低自由生成带来的幻觉问题。
- **【工具增强校验】提升可信度**：基于高德 API 获取 POI、酒店、餐饮和天气候选，结合 Pydantic schema、业务规则校验、预算复算、失败重试和 fallback，提升输出稳定性。
- **【多候选 Rerank】选择更优方案**：支持对多个候选行程按预算贴合、POI grounding、餐饮多样性等指标打分选优，并构建 SFT / DPO / Best-of-N 评测流程。
- **【RAG 扩展设计】增强可执行性**：设计将景点开放时间、预约规则、游玩时长、人群约束和雨天备选等知识注入 PlannerContext 的方案，增强行程落地可执行性。

开源项目

CommandBox	常用命令管理工具	https://github.com/GotYoo/CommandBox
------------	----------	---

项目背景：开发和实验过程中常用命令分散在终端历史、笔记和聊天记录中，存在查找慢、复制易错、上下文缺失等问题。

- **【产品方案】降低复用成本**：参考 Raycast、VSCode Command Palette 等产品，将命令管理抽象为“搜索优先 + 分类归档 + 卡片详情”的桌面工具。
- **【桌面端实现】核心功能实现**：使用 Python + PyQt5 开发 Windows 桌面应用，支持分类、搜索、详情编辑、复制、导入导出和本地 JSON 数据持久化。
- **【体验优化】围绕高频复制**：设计双击复制、代码块复制图标、状态反馈和可直接编辑详情区，并重构 UI 以提升信息密度和可读性。