



吴高

算法工程师 | Agent / 大模型后训练

156-7170-6776 | wu\_gao@foxmail.com | 浙江大学硕士 (2027 届)



## 教育经历

浙江大学 | 控制科学与工程学院 | 控制科学与工程 (硕士)

2024.09 - 2027.06

科研成果: GUI Agent 方向在投/公开论文 5 篇 (第一/共同一作 3 篇、二作 1 篇、共同作者 1 篇)。

西北工业大学 | 自动化学院 | 自动化 (本科, 保研)

2020.09 - 2024.06

GPA 3.73/4.1, 专业排名 11/178; 校一等奖学金、校优秀毕业生、校优秀大学生。

## 实习经历

阿里巴巴通义实验室 | AI 研发实习生

2026.06 - 至今

- 面向 Qwen 系列模型后训练与评测中的大规模 Agent Rollout 需求, 参与建设统一任务编排与运行基础设施, 贯通任务定义、资源调度、环境执行、过程观测、产物回收及结果评估链路。
- 负责 Rollout 任务生命周期与运行环境建设, 标准化任务模板、参数配置、容器镜像、运行目录及评测产物协议; 完善环境初始化、健康检查、模型调用、异常诊断与失败恢复链路。
- 平台支持 100+ 个任务模板、4 万级并发与百万级日处理量, 为大模型后训练提供可扩展的在线 Rollout 基础设施。

快手主站技术部 | AI 研究实习生

2025.08 - 2026.06

- 参与企业级 Android 通用智能回测平台建设, 负责 GUI Agent 与真实测试业务结合, 推动自然语言任务理解、自动执行、结果验证与知识沉淀落地。
- 参与平台方案设计与 Agent 核心链路实现, 抽象跨场景测试流程并优化结果分析, 降低测试用例编写和维护成本。
- 完整任务端到端执行成功率达到 90%; Agent 正/负例与人工标注 GT 对齐后的判定准确率达到 95%。

## 科研与项目经历

MobileForge: 闭环自进化移动端 GUI Agent 框架 | 共同第一作者

论文 / 代码 / 数据 / 模型

- 共同设计并实现闭环自进化后训练框架, 构建从目标应用探索、课程任务挖掘到策略优化的完整链路, 解决任务、轨迹和奖励标注成本高的问题。
- 聚焦交互采样与层级反馈评测, 将任务级/步骤级反馈及失败提示转化为 step-level GRPO 数据, 基于 VeRL 完成策略优化。
- Qwen3-VL-8B 在 AndroidWorld Pass@3 从 55.2% 提升至 67.2%; ForgeOwl-8B 达到 77.6%, MobileWorld 纯 GUI 任务成功率由 37.6% 提升至 41.0%。

带噪交互场景下的 GUI Agent 经验记忆自进化系统 | 项目负责人

第一作者在投

- 面向用户指令带噪、跨应用证据链缺失及反复点击/幻觉等失败模式, 主导建设 GUI Agent 经验记忆自进化系统, 提升开放移动端任务的鲁棒性。
- 主导系统设计、实现与评测, 完成轨迹解析、经验抽取/晋升、质量门控、向量检索、LLM 重排及提示注入等核心模块, 将失败轨迹沉淀为操作流程、风险规则与跨应用 workflow。
- 在 117 个 MobileWorld 任务上, Seed Agent Pass@5 由 79/117 (67.5%) 提升至 91/117 (77.8%), 并沉淀跨应用经验库。

MemGUI: 长程 GUI Agent 记忆评测与主动上下文管理 | 第二作者

论文 / 代码 / 数据 / 模型

- 参与主动上下文管理与训练评测链路建设; 以 Context-as-Action 统一生成 GUI 操作和记忆管理动作, 使模型自主决定何时压缩历史、写入/更新记忆并保留关键信息。
- 从 128 个种子任务扩展任务并采集、过滤教师模型轨迹, 构建含 2,956 条轨迹、64,430 个有效步骤的 MemGUI-3K 数据集, 基于 Qwen3-VL-8B 完成 LoRA SFT。
- MemGUI-Agent-235B 在 MemGUI-Bench 达到 37.5% Pass@1 / 62.5% Pass@3; 8B-SFT 的 Pass@1 由 9.4% 提升至 23.4%, 并在域外 MobileWorld 提升 8.5 个百分点。

招商银行金融 Text-to-SQL 模型训练与推理优化 | 项目负责人

第 6/18

- 在两天内完成金融场景数据构建、模型训练与昇腾环境推理优化; 设计强模型生成 SQL、PostgreSQL 只读执行验真、结果比对和失败反馈修复流程, 构建 1,558 个 episode 的训练集。
- 基于 Qwen3.5-2B 搭建 LoRA SFT、GRPO 与执行结果奖励训练闭环, 验证集准确率由 38.05% 提升至 75.58%, 推理吞吐由 0.33 提升至 2.48 req/s。

## 技术栈与奖项

技术栈: Python、PyTorch、VeRL、ms-swift、LoRA SFT、GRPO、Android 自动化、PostgreSQL

竞赛奖项: 全国大学生数学竞赛 (非数学类) 一等奖 | 美国大学生数学建模竞赛 (MCM) M 奖