

谢睿

电话: 18136603723

邮箱: sharryXR@sjtu.edu.cn

学校: 上海交通大学 计算机学院

学位: 计算机科学与技术 直博一年级 导师: 陈露



教育背景

上海交通大学

2025.09 – 至今

计算机科学与技术 直博 导师: 陈露 GPA: 3.89 / 4.0

上海交通大学

2021.09 – 2025.06

计算机科学与技术 工学学士 计算机学院

核心学积分: 91.70/100 排名: 23/103 突出课程: 离散数学 (100)、编译原理 (100)、计算机系统结构 (99)、操作系统 (98)、数据结构 (98)、概率统计 (99)

论文发表

[1] **GUIDE: Resolving Domain Bias in GUI Agents through Real-Time Web Video Retrieval and Plug-and-Play Annotation**

Rui Xie, Zhi Gao, Chenrui Shi, Zirui Shang, Lu Chen*, Qing Li*

[ECCV 2026 Under Review](#) | 第一作者

[2] **MatToolBench: A Real-Environment Benchmark for Evaluating Multimodal Agents on Professional Materials Science Software**

Mei Wu[†], Rui Xie[†], Runyu Zhang, Lu Chen*, Bo Chen, Kai Yu*, Xin Chen* († 共同一作)

[In Preparation \(Target: WAICA 2026\)](#) | 共同第一作者

[3] **TongUI: Internet-Scale Trajectories from Multimodal Web Tutorials for Generalized GUI Agents**

Bofei Zhang[†], Zirui Shang[†], Zhi Gao[†], Wang Zhang, Rui Xie, Xiaojian Ma, Tao Yuan, Xinxiao Wu, Song-Chun Zhu, Qing Li*

[AAAI 2026](#) | 参与作者

[4] **Mobile-Env: Building Qualified Evaluation Benchmarks for LLM-GUI Interaction**

Danyang Zhang, Zhennan Shen, Rui Xie, Situo Zhang, Tianbao Xie, Zihan Zhao, Siyuan Chen, Lu Chen, Hongshen Xu, Ruisheng Cao, Kai Yu

[计算机学报 Under Review](#) | 第三作者

研究经历

GUIDE: 基于视频检索增强的 GUI Agent 领域偏差消除框架

2025.03 – 至今

上海交通大学 X-LANCE Lab / 北京通用人工智能研究院 (BIGAI)

第一作者

- 提出 **GUIDE** 框架, 针对 GUI Agent 在特定领域软件操作中的规划 (**planning**) 与定位 (**grounding**) 偏差问题, 设计了一种无需训练的、即插即用的解决方案。
- 设计了基于字幕的 **Video-RAG** 检索管道, 通过三阶段渐进式检索 (领域分类、主题提取、相关性匹配) 从网络教程视频中精准定位任务相关内容。构建了基于逆动力学范式的全自动标注管道, 利用 VLM 从增强 UI

元素检测的连续关键帧推断帧间动作，生成可迁移的规划与定位知识。

- 在 OSWorld 基准上验证，作为即插即用组件集成至多智能体系统和单模型 Agent 均**一致提升 5%+** 成功率并减少执行步数。

MatToolBench: 材料科学领域多模态智能体真实环境基准测试 2025.10 – 至今
上海交通大学 X-LANCE Lab / 苏州实验室 共同第一作者

- 主导构建了首个面向材料科学专业软件的**真实环境 GUI Agent 基准**，涵盖 10 种工具、204 个任务，覆盖 GUI 操作、代码执行和跨工具交互三种任务模态。
- 设计并实现了基于 Windows 虚拟机的**隔离测试环境**，支持 Docker 容器化部署及 Azure ML 并行评测，实现了标准化的环境重置与 Gym 接口适配。提出了**多层级评测框架**：文本识别、视觉-文本混合、LLM 美学评判、代码匹配四种评估方式，适配不同任务特性。
- 对多种 SOTA 多模态 Agent 进行系统性性能基准测试，揭示通用 Agent 在专业领域的显著性能差距。

TongUI: 基于网络多模态教程的大规模 GUI Agent 数据构建 2024.10 – 2025.04
北京通用人工智能研究院 (BIGAI) / 上海交通大学 参与作者

- 参与构建了 **GUI-Net-1M**——开源 GUI 轨迹数据集（100 万条轨迹，5 种操作系统，280+ 应用）。
- 负责部分 GUI 基准测试的**实验测试**工作，主要包括 ScreenSpot 离线评测及 MiniWob 在线评测。
- 参与基于 Qwen2.5-VL 的 **SFT 训练**，开发 TongUI Agent，在多个基准上超越基线 10%+。

MacOS-Agent: 面向 macOS 的多应用 GUI Agent 系统 2025.03 – 2025.05
北京通用人工智能研究院 (BIGAI) 研究实习

- 参与开发并测试部署于 macOS 系统的多应用 GUI Agent，支持 10 种桌面应用的常用任务自动化执行与跨应用协同。
- 主要负责 **Browser**、**Calendar**、**WeChat** 三个应用 Agent 的开发，实现了网页浏览与信息检索收集、日历管理、消息发送等跨应用联动任务。

GUI Agent 强化学习训练 2025.06 – 2025.08
北京通用人工智能研究院 (BIGAI) 研究实习

- 参与 GUI Agent 多轮强化学习训练的早期工作，负责 **RL 训练任务生成**模块的设计与实现，为强化学习训练构建多样化的 GUI 任务集。
- 将训练框架中的强化学习方法从 **GRPO** 替换为 **DAPO**，并基于此进行了 GUI Agent 的强化学习训练。

Mobile-Env: LLM-GUI 交互评测基准构建 2024.04 – 2024.11
上海交通大学 X-LANCE Lab 第三作者

- 在 Android 移动环境中为 LLM GUI Agent 构建评测基准，设计并编写饿了么 App 测试任务集，完成流量爬取与本地隔离环境部署。
- 对多种主流大模型进行 GUI Agent 性能基准测试，引入思维链规范智能体行为，提升图标点击准确率。

专业能力

- 研究方向**：大模型智能体 (LLM/VLM Agent)，尤其在 **GUI Agent** 领域有持续深入的研究经验，覆盖 Agent 框架设计、大规模轨迹数据构建与模型微调、强化学习训练、基准构建与评测等多个环节。
- 科研能力**：具备 VLM 模型 SFT/LoRA 微调及**强化学习训练**经验（熟悉 PPO、GRPO、DAPO 等方法及其在 GUI Agent 中的应用）；具有跨平台（Android/Windows/macOS/Web）GUI Agent 系统的开发与评测经验。
- 语言**：CET-6 564

荣誉奖项