Novelty Assessment Report

Paper: ScaleCUA: Scaling Open-Source Computer Use Agents with Cross-Platform Data
PDF URL: https://openreview.net/pdf?id=yBFUqdJFZn
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-29

Abstract

Vision-Language Models (VLMs) have enabled computer use agents (CUAs) that operate GUIs autonomously, showing great potential, yet progress is limited by the lack of large-scale, open-source computer use data and foundation models. In this work, we introduce ScaleCUA, a step toward scaling open-source CUAs. It offers a large-scale dataset spanning 6 operating systems and 3 task domains, built via a closed-loop pipeline uniting automated agents with human experts. Trained on this scaled-up data, ScaleCUA can operate seamlessly across platforms. Specifically, it delivers strong gains over baselines (+26.6 on WebArena-Lite-v2, +10.7 on ScreenSpot-Pro) and sets new state-of-the-art results (94.4% on MMBench-GUI L1-Hard, 60.6% on OSWorld-G, 47.4% on WebArena-Lite-v2). These findings underscore the power of data-driven scaling for general-purpose computer use agents. We will release data, models, and code to advance future research.

Core Task Landscape

This paper addresses: **Scaling Open-Source Computer Use Agents with Cross-Platform Data**

A total of **13 papers** were analyzed and organized into a taxonomy with **10 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Foundation Models and Architectures for GUI Agents**
- **Data Collection and Scaling Pipelines**
- **Evaluation Frameworks and Benchmarks**
- **Specialized Agent Capabilities and Applications**

Complete Taxonomy Tree

- Scaling Open-Source Computer Use Agents with Cross-Platform Data Survey Taxonomy
- Foundation Models and Architectures for GUI Agents
  - Open-Source Foundation Action Models (3 papers)
  - [1] Os-atlas: A foundation action model for generalist gui agents (Wu, 2024) View paper
  - [2] Opencua: Open foundations for computer-use agents (Wang Xin-yuan, 2025) View paper
  - [8] Mobile-agent-v3: Fundamental agents for gui automation (Ye, 2025) View paper
  - Hybrid Action Integration (1 papers)
  - [6] UltraCUA: A Foundation Model for Computer Use Agents with Hybrid Action (Yang, 2025) View paper
  - Unified Cross-Environment Architectures (2 papers)
  - [11] OmniActor: A Generalist GUI and Embodied Agent for 2D&3D Worlds (Zeng Zhixiong, 2025) View paper
  - [12] Surfer 2: The Next Generation of Cross-Platform Computer Use Agents (Andreux, 2025) View paper
- Data Collection and Scaling Pipelines
  - Cross-Platform Dataset Construction ★ (2 papers)
  - [0] ScaleCUA: Scaling Open-Source Computer Use Agents with Cross-Platform Data (Anon et al., 2026) View paper
  - [5] CCAgent: Coordinating Collaborative Data Scaling for Operating System Agents via Web3 (Liang Chen, 2025) View paper
  - Mobile-Specific Annotation Datasets (1 papers)
  - [7] Amex: Android multi-annotation expo dataset for mobile gui agents (Yuxiang Chai, 2025) View paper
- Evaluation Frameworks and Benchmarks
  - Hierarchical Multi-Platform Benchmarks (1 papers)
  - [3] Mmbench-gui: Hierarchical multi-platform evaluation framework for gui agents (Wang Xuehui, 2025) View paper
  - Fine-Grained State Control Evaluation (1 papers)
  - [10] FineState-Bench: A Comprehensive Benchmark for Fine-Grained State Control in GUI Agents (JI Fengxian, 2025) View paper
  - Step-Level Critic Model Frameworks (1 papers)
  - [9] OS-Oracle: A Comprehensive Framework for Cross-Platform GUI Critic Models (Zhenyu Wu, 2025) View paper
- Specialized Agent Capabilities and Applications
  - Cognitive Resource Management for Long-Horizon Tasks (1 papers)
  - [13] Don't Lose the Thread: Empowering Long-Horizon LLM Agents with Cognitive Resource Self-Allocation (Z Zhu, n.d.) View paper
  - Cross-Platform Security Applications (1 papers)
  - [4] Towards Cross-platform Detection of Cyberattacks Using Micro-agents. (Xinxing Zhao, 2019) View paper

Narrative

Core task: Scaling open-source computer use agents with cross-platform data. The field of computer use agents has rapidly evolved around four main branches. Foundation Models and Architectures for GUI Agents explores the underlying neural architectures and pretraining strategies that enable agents to perceive and interact with graphical interfaces. Data Collection and Scaling Pipelines focuses on methods for gathering, annotating, and synthesizing large-scale interaction traces across diverse operating systems and applications, addressing the bottleneck of training data availability. Evaluation Frameworks and Benchmarks develops standardized testbeds and metrics to measure agent performance on realistic tasks, such as MMBench-GUI[3] and FineState-Bench[10]. Specialized Agent Capabilities and Applications examines domain-specific skills—ranging from mobile navigation in Mobile-Agent-v3[8] to web browsing in Surfer 2[12]—and how agents generalize across platforms.

A particularly active line of work centers on cross-platform dataset construction, where researchers aim to unify interaction data from Windows, macOS, Linux, mobile, and web environments to improve agent robustness. ScaleCUA[0] exemplifies this direction by systematically aggregating diverse platform traces to train open-source models at scale. Closely related efforts include OS-Atlas[1], which emphasizes operating-system-level grounding, and CCAgent[5], which tackles cross-platform consistency in action spaces. Meanwhile, OpenCUA[2] and UltraCUA[6] explore complementary strategies for data synthesis and quality filtering. The main trade-off in this cluster revolves around breadth versus depth: some works prioritize coverage across many platforms, while others focus on high-fidelity annotations within a narrower scope. ScaleCUA[0] sits squarely in the breadth-focused camp, leveraging cross-platform diversity to enhance generalization, whereas CCAgent[5] places greater emphasis on ensuring semantic alignment of actions across different GUI paradigms.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. CCAgent: Coordinating Collaborative Data Scaling for Operating System Agents via Web3

Authors: Liang Chen, HaoZhe Zhao, Yinzhen Huang, Yang Luo, Tsekai Lin, et al. (11 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

The current AI revolution, fueled by Large Language Models (LLMs), heavily relies on vast open-access internet data. However, the Operating System (OS) Agent field faces a significant data sparsity challenge due to the lack of public data collection systems and privacy concerns. To address this, we introduce CCAgent Net, a system that coordinates and incentivizes internet users to contribute to scaling OS agent datasets. Furthermore, we propose GUI-Pipe, an automated data post-processing pipelin...

Relationship Analysis

Both papers belong to the Cross-Platform Dataset Construction category, focusing on large-scale data collection for computer use agents across multiple operating systems. They overlap in addressing the data scarcity challenge through automated pipelines and multi-platform coverage (desktop, mobile, web), with both producing large instruction-based GUI datasets. However, ScaleCUA emphasizes a dual-loop pipeline combining automated agents with human experts for quality assurance, while CCAgent introduces a Web3-based crowdsourcing system (CCAgent Net) that coordinates and incentivizes internet users to contribute data, representing fundamentally different approaches to scaling data collection.

Contributions Analysis

Overall novelty summary. The paper contributes a large-scale cross-platform dataset spanning six operating systems and three task domains, alongside a family of foundation models (ScaleCUA) trained on this data. Within the taxonomy, it resides in the 'Cross-Platform Dataset Construction' leaf under 'Data Collection and Scaling Pipelines,' sharing this leaf with only one sibling paper (OS-Atlas). This positioning indicates a relatively sparse research direction focused specifically on multi-OS data aggregation, distinguishing it from mobile-only annotation efforts and from foundation model architectures that consume such data.

The taxonomy reveals neighboring work in 'Open-Source Foundation Action Models' (three papers on VLM-based GUI agents) and 'Mobile-Specific Annotation Datasets' (one paper on mobile GUI annotations). The paper bridges data collection and model training, connecting to both the dataset construction branch and the foundation model branch. Its scope explicitly includes automated data pipelines with human-in-the-loop validation, aligning with the leaf's scope note emphasizing 'automated or hybrid collection pipelines.' The exclude note clarifies that mobile-only datasets belong elsewhere, reinforcing this work's cross-platform emphasis.

Among 29 candidates examined, the data pipeline contribution (Contribution A) identified one refutable candidate from 10 examined, while the base model contribution (Contribution B) also found one refutable candidate from 10 examined. The evaluation contribution (Contribution C) showed no clear refutation across nine candidates. These statistics suggest that within the limited search scope, the data pipeline and model architecture face some prior overlap, whereas the evaluation insights appear more distinctive. The relatively low refutation counts (one per contribution) indicate that most examined candidates address different facets or scales of the problem.

Given the limited search scope of 29 candidates from top-K semantic matches, this analysis captures the most semantically proximate prior work but does not constitute an exhaustive field survey. The sparse taxonomy leaf (two papers total) and low refutation rates suggest the work occupies a relatively underexplored niche within computer use agents, though the presence of any refutable candidates indicates incremental overlap with existing cross-platform data efforts. The evaluation contribution appears most novel within this constrained search.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Cross-platform interactive data pipeline for computer use agents

Description: The authors propose a dual-loop data acquisition pipeline that combines automated agent exploration with human expert supervision to collect computer-use data across six operating systems (Windows, macOS, Linux, Android, iOS, Web) and three task domains (understanding, grounding, task completion). This pipeline addresses the scarcity of computer-use training data by balancing automation with quality control.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Chouette: An Automated Cross-Platform UI Crawler for Improving App Quality

URL: View paper

Brief Assessment

Chouette[39] is a GUI crawler for automated testing at Duolingo, not a data collection pipeline for training computer use agents. It focuses on bug detection through computer vision-based exploration, whereas the original paper describes a dual-loop system combining automated agents with human supervision to collect training data across operating systems and task domains.

---

2. A Multi-Agent Monitoring System for Computer Networks

URL: View paper

Brief Assessment

Multi-Agent Network Monitoring[40] focuses on network device monitoring using SNMP agents and Apache Kafka, not on cross-platform GUI data collection for computer use agents. The domains are fundamentally different.

---

3. Exploring the Integration of Generative AI Tools in Software Testing Education: A Case Study on ChatGPT and Copilot for Preparatory Testing Artifacts in Postgraduate â¦

URL: View paper

Brief Assessment

GenAI Testing Education[33] focuses on integrating generative AI tools (ChatGPT and Copilot) in software testing education for postgraduate students. It does not address cross-platform computer use dataset collection, automated agent exploration, or human supervision for GUI interaction data.

---

4. Opencua: Open foundations for computer-use agents

URL: View paper

Prior Art Analysis

OpenCUA[2] demonstrates that a similar dual-loop data acquisition pipeline combining automated agents with human supervision for cross-platform computer-use data collection was already implemented. The candidate paper describes an annotation infrastructure that captures human demonstrations across multiple operating systems (Windows, macOS, Ubuntu) with both automated and human-in-the-loop collection, directly paralleling the original paper's claimed contribution. Both papers employ similar strategies of combining automated exploration with human expert curation to balance data quality and scale.

Evidence

Evidence 1 - **Rationale**: Both describe dual mechanisms for data collection. The original's 'agent-environment interaction loop' and 'agent-human hybrid data acquisition loop' are functionally equivalent to the candidate's system that 'runs on annotators' personal computers' while capturing comprehensive interaction data. - **Original**: the agent-environment interaction loop enables automated agents to interact with diverse gui environments, while the agent-human hybrid data acquisition loop integrates expert-collected trajectories to ensure coverage and quality - **Candidate**: To address this, we developed a user-friendly annotation tool that streamlines the collection and verification of computer-use demonstrations (figure 3), runs on annotators' personal computers and records demonstrations in the background, capturing: (1) screen videos, (2) mouse and keyboard signals,...

---

5. 3EED: Ground everything everywhere in 3D

URL: View paper

Brief Assessment

3EED[32] focuses on 3D visual grounding for embodied robots across vehicle/drone/quadruped platforms in outdoor environments, not on computer use agents or GUI interaction across operating systems.

---

6. Multi-Agent Systems in AIOps: Enhancing Detection, Diagnosis, and Remediation

URL: View paper

Brief Assessment

Multi-Agent AIOps[35] focuses on IT operations management using multi-agent systems for incident detection and remediation in enterprise environments, not on cross-platform GUI data collection or computer use agent training pipelines.

---

7. Omniact: A dataset and benchmark for enabling multimodal generalist autonomous agents for desktop and web

URL: View paper

Brief Assessment

OmniAct[36] focuses on desktop and web task automation using PyAutoGUI scripts with human annotation, not on cross-platform data collection pipelines combining automated agents with human supervision across six operating systems.

---

8. REPP: A robust cross-platform solution for online sensorimotor synchronization experiments

URL: View paper

Brief Assessment

REPP[38] focuses on sensorimotor synchronization experiments for rhythm/tapping tasks in psychology research, not on computer use agents or GUI automation across operating systems.

---

9. Crab: Cross-platfrom agent benchmark for multi-modal embodied language model agents

URL: View paper

Brief Assessment

Crab[34] focuses on cross-environment agent benchmarking and evaluation frameworks, not on data collection pipelines. The original paper's contribution is specifically about dual-loop data acquisition combining automated agents with human supervision for training data collection, which is not addressed in Crab[34].

---

10. Automated Database Tuning Using AI: A Comprehensive Framework for Real-Time Performance Optimization

URL: View paper

Brief Assessment

AI Database Tuning[37] focuses on database performance optimization using AI techniques, not on cross-platform computer use data collection or GUI agent training pipelines.

Contribution 2: ScaleCUA family of base agent models with unified action space

Description: The authors develop a series of vision-language models (3B, 7B, 32B parameters) built on Qwen2.5-VL that support three inference paradigms: grounding mode for UI element localization, direct action mode for efficient task completion, and reasoned action mode with chain-of-thought reasoning. The models use a unified action space enabling seamless cross-platform interaction.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Aria-ui: Visual grounding for gui instructions

URL: View paper

Brief Assessment

Aria-UI[21] focuses on visual grounding for GUI instructions as a specialized grounding model, not on developing a family of base agent models with unified action spaces across platforms. The candidate addresses grounding specifically, while the original contribution encompasses broader agent model architecture.

---

2. CogAgent: A Visual Language Model for GUI Agents

URL: View paper

Brief Assessment

CogAgent[15] focuses on GUI understanding through dual-resolution visual encoding (low-res 224×224 and high-res 1120×1120) but does not present a unified action space framework spanning multiple platforms. The candidate addresses visual perception challenges rather than cross-platform action unification.

---

3. Aguvis: Unified pure vision agents for autonomous gui interaction

URL: View paper

Prior Art Analysis

Aguvis[16] demonstrates prior work on unified vision-based GUI agents with standardized action spaces across platforms. The candidate paper presents a framework that operates purely through visual observations with a standardized pyautogui-based action space, enabling cross-platform interaction across desktop, mobile, and web environments. This directly addresses the same core innovation claimed in the original paper: unifying perception, reasoning, and action into vision-language models with consistent cross-platform action representations.

Evidence

Evidence 1 - **Rationale**: Both papers describe unified action spaces with multiple inference modes (grounding, direct action, reasoned action in the original; similar capabilities in the candidate), showing prior work on this specific technical approach. - **Original**: we design a unified action space, allowing for more consistent and efficient interaction with diverse real-world environments... building upon this corpus, we train a series of base agent models termed asscalecua with qwen2.5vl (bai et al., 2025). it supports three inference paradigms to offer enhan... - **Candidate**: for action execution, we adopt pyautoguias our universal interaction interface, supplemented by a flexible plugin system. the pyautogui library provides a comprehensive set of programmatic commands that mirror human input behaviors, allowing us to represent gui interactions consistently across platf...

Evidence 2 - **Rationale**: Both papers detail cross-platform action spaces combining universal and platform-specific operations, demonstrating prior work on this unified action representation approach. - **Original**: table 14 summarizes our cross-platform action space spanning desktop, mobile, and web. it combines universal operations (e.g., click, write, etc.) with platform-specific actions (e.g., long press and open appfor mobile), ensuring consistent behavior modeling and simplifying downstream policy learnin... - **Candidate**: as shown in table 9, this standardized action space enables the model to translate its inner monologue into concrete actions without requiring environment-specific design. our plugin system extends the base pyautogui action space to handle platform-specific requirements while maintaining the natural...

---

4. ScreenAgent: A Vision Language Model-driven Computer Control Agent

URL: View paper

Brief Assessment

ScreenAgent[22] focuses on a different architecture and training approach. While both involve vision-language models for GUI control, ScreenAgent[22] uses a control pipeline with planning/acting/reflecting phases and trains on a different dataset (ScreenAgent dataset vs. ScaleCUA's cross-platform data). The unified action space and three inference paradigms (grounding, direct action, reasoned action) described in the original contribution are not demonstrated in ScreenAgent[22].

---

5. Mobilevlm: A vision-language model for better intra-and inter-ui understanding

URL: View paper

Brief Assessment

MobileVLM[18] focuses on Chinese mobile UI understanding with intra- and inter-UI pre-training tasks, not on cross-platform computer use agents with unified action spaces spanning desktop, mobile, and web environments as described in the original contribution.

---

6. Seed1. 5-vl technical report

URL: View paper

Brief Assessment

Seed1.5-VL[20] focuses on general-purpose multimodal understanding and reasoning across diverse tasks, not specifically on developing a family of GUI agent models with unified action spaces for cross-platform interaction. The architectural and training approaches differ fundamentally from ScaleCUA's agent-centric design.

---

7. Gui-r1: A generalist r1-style vision-language action model for gui agents

URL: View paper

Brief Assessment

GUI-R1[19] focuses on reinforcement fine-tuning (RFT) for GUI agents using a unified action space reward function, rather than developing a family of vision-language models with multiple inference paradigms (grounding, direct action, reasoned action) as described in the original contribution.

---

8. Mug: Interactive multimodal grounding on user interfaces

URL: View paper

Brief Assessment

Mug[17] focuses on interactive multimodal grounding for single-screen UI tasks with user-agent collaboration, not on building general-purpose cross-platform computer use agents with unified action spaces across multiple operating systems.

---

9. Infigui-r1: Advancing multimodal gui agents from reactive actors to deliberative reasoners

URL: View paper

Brief Assessment

InfiGUI-R1[23] focuses on reasoning-centric training methodology for GUI agents rather than unified action space design. The candidate emphasizes transforming agents from reactive actors to deliberative reasoners through spatial reasoning distillation and reinforcement learning, which is a different technical approach from the original paper's focus on scaling data and unified action spaces across platforms.

---

10. Internvl3. 5: Advancing open-source multimodal models in versatility, reasoning, and efficiency

URL: View paper

Brief Assessment

InternVL3.5[14] focuses on general-purpose multimodal models with capabilities across text, reasoning, and agentic tasks, but does not specifically propose a unified action space for GUI interaction across platforms as the core contribution. The candidate addresses GUI tasks as one application among many, rather than specializing in cross-platform computer use agents with unified action paradigms.

Contribution 3: Comprehensive evaluation and insights for computer use agents

Description: The authors perform extensive experiments across GUI understanding, grounding, and task completion benchmarks on multiple platforms. Their evaluation provides fundamental insights into factors affecting agent performance, including data scaling effects, resolution trade-offs, inference mode comparisons, and cross-platform training strategies, establishing new state-of-the-art results on several benchmarks.

This contribution was assessed against **9 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Aria-ui: Visual grounding for gui instructions

URL: View paper

Brief Assessment

While Aria-UI[21] conducts evaluations on GUI grounding benchmarks, it does not provide the comprehensive cross-platform evaluation spanning understanding, grounding, and task completion with insights into data scaling, resolution trade-offs, and cross-platform training strategies that characterize the original contribution.

---

2. Navigating the digital world as humans do: Universal visual grounding for gui agents

URL: View paper

Brief Assessment

Universal Visual Grounding[25] focuses on visual grounding models for GUI agents across platforms, not comprehensive evaluation frameworks. The candidate evaluates grounding accuracy and agent performance but does not provide the systematic cross-platform evaluation infrastructure, data scaling insights, or resolution/inference mode trade-offs that characterize the original paper's contribution.

---

3. Gta1: Gui test-time scaling agent

URL: View paper

Brief Assessment

GTA1[29] focuses specifically on GUI grounding and test-time scaling strategies for action proposal selection, rather than providing comprehensive cross-platform evaluation insights across understanding, grounding, and task completion benchmarks as claimed in the original contribution.

---

4. Aguvis: Unified pure vision agents for autonomous gui interaction

URL: View paper

Brief Assessment

[Final Audit Failure] The model insisted on a refutation claim but failed to provide verifiable evidence after multiple retries. Marked as cannot_refute for safety. Please manually verify the candidate text.

---

5. Gui agents with foundation models: A comprehensive survey

URL: View paper

Brief Assessment

GUI Agents Foundation[30] is a survey paper that reviews existing work on GUI agents rather than conducting original evaluations. It does not present novel experimental results or insights that would refute the original paper's claim to comprehensive evaluation across platforms.

---

6. Scaling Computer-Use Grounding via User Interface Decomposition and Synthesis

URL: View paper

Brief Assessment

UI Decomposition Synthesis[26] focuses on GUI grounding methodology and dataset construction (JEDI, OSWorld-G), not comprehensive cross-platform agent evaluation. The candidate does not demonstrate prior work on the original paper's multi-platform evaluation framework covering understanding, grounding, and task completion with insights on data scaling, resolution trade-offs, and cross-platform training strategies.

---

7. UItron: Foundational GUI Agent with Advanced Perception and Planning

URL: View paper

Brief Assessment

UItron[27] focuses on building a foundational GUI agent with systematic data engineering and interactive infrastructure, rather than providing comprehensive evaluation insights across platforms. The candidate emphasizes training paradigms and Chinese app scenarios, not fundamental insights into factors affecting agent performance like the original paper's focus on data scaling effects, resolution trade-offs, and cross-platform training strategies.

---

8. Seeclick: Harnessing gui grounding for advanced visual gui agents

URL: View paper

Brief Assessment

SeeClick[24] focuses on GUI grounding for visual agents on mobile/desktop/web, not comprehensive cross-platform evaluation of understanding, grounding, and task completion with insights into data scaling, resolution trade-offs, and training strategies as claimed in the original work.

---

9. Gui agents: A survey

URL: View paper

Brief Assessment

GUI Agents Survey[31] is a survey paper that reviews existing benchmarks and evaluation metrics across the field, rather than presenting original experimental evaluations. It does not demonstrate that similar comprehensive evaluation work existed prior to the original paper's contributions.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] ScaleCUA: Scaling Open-Source Computer Use Agents with Cross-Platform Data View paper
- [1] Os-atlas: A foundation action model for generalist gui agents View paper
- [2] Opencua: Open foundations for computer-use agents View paper
- [3] Mmbench-gui: Hierarchical multi-platform evaluation framework for gui agents View paper
- [4] Towards Cross-platform Detection of Cyberattacks Using Micro-agents. View paper
- [5] CCAgent: Coordinating Collaborative Data Scaling for Operating System Agents via Web3 View paper
- [6] UltraCUA: A Foundation Model for Computer Use Agents with Hybrid Action View paper
- [7] Amex: Android multi-annotation expo dataset for mobile gui agents View paper
- [8] Mobile-agent-v3: Fundamental agents for gui automation View paper
- [9] OS-Oracle: A Comprehensive Framework for Cross-Platform GUI Critic Models View paper
- [10] FineState-Bench: A Comprehensive Benchmark for Fine-Grained State Control in GUI Agents View paper
- [11] OmniActor: A Generalist GUI and Embodied Agent for 2D&3D Worlds View paper
- [12] Surfer 2: The Next Generation of Cross-Platform Computer Use Agents View paper
- [13] Don't Lose the Thread: Empowering Long-Horizon LLM Agents with Cognitive Resource Self-Allocation View paper
- [14] Internvl3. 5: Advancing open-source multimodal models in versatility, reasoning, and efficiency View paper
- [15] CogAgent: A Visual Language Model for GUI Agents View paper
- [16] Aguvis: Unified pure vision agents for autonomous gui interaction View paper
- [17] Mug: Interactive multimodal grounding on user interfaces View paper
- [18] Mobilevlm: A vision-language model for better intra-and inter-ui understanding View paper
- [19] Gui-r1: A generalist r1-style vision-language action model for gui agents View paper
- [20] Seed1. 5-vl technical report View paper
- [21] Aria-ui: Visual grounding for gui instructions View paper
- [22] ScreenAgent: A Vision Language Model-driven Computer Control Agent View paper
- [23] Infigui-r1: Advancing multimodal gui agents from reactive actors to deliberative reasoners View paper
- [24] Seeclick: Harnessing gui grounding for advanced visual gui agents View paper
- [25] Navigating the digital world as humans do: Universal visual grounding for gui agents View paper
- [26] Scaling Computer-Use Grounding via User Interface Decomposition and Synthesis View paper
- [27] UItron: Foundational GUI Agent with Advanced Perception and Planning View paper
- [28] Large language model-brained gui agents: A survey View paper
- [29] Gta1: Gui test-time scaling agent View paper
- [30] Gui agents with foundation models: A comprehensive survey View paper
- [31] Gui agents: A survey View paper
- [32] 3EED: Ground everything everywhere in 3D View paper
- [33] Exploring the Integration of Generative AI Tools in Software Testing Education: A Case Study on ChatGPT and Copilot for Preparatory Testing Artifacts in Postgraduate â¦ View paper
- [34] Crab: Cross-platfrom agent benchmark for multi-modal embodied language model agents View paper
- [35] Multi-Agent Systems in AIOps: Enhancing Detection, Diagnosis, and Remediation View paper
- [36] Omniact: A dataset and benchmark for enabling multimodal generalist autonomous agents for desktop and web View paper
- [37] Automated Database Tuning Using AI: A Comprehensive Framework for Real-Time Performance Optimization View paper
- [38] REPP: A robust cross-platform solution for online sensorimotor synchronization experiments View paper
- [39] Chouette: An Automated Cross-Platform UI Crawler for Improving App Quality View paper
- [40] A Multi-Agent Monitoring System for Computer Networks View paper