

OpenGCUA releases first open stack for generalist computer-use agents

Source copy kept by wheresthe.ai for the story "OpenGCUA releases first open stack for generalist computer-use agents" (<https://wheresthe.ai/d/intl-opengcua-releases-first-open-stack-for-generalist-compu-2026-10/>).

Source: <https://arxiv.org/abs/2610.22200>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: 10ec4e83a474773fa7f5cb0fa21c8ba4567e912cd1d5603542ba670e974fb0f9

-->

Computer Science > Computer Vision and Pattern Recognition

arXiv:2610.22200 (cs)

[Submitted on 15 Oct 2026]

Title: OpenGCUA: Toward a Fully Open Stack for Generalist Computer Use

Authors: Hanzhang Zhou , Long Li , Quyu Kong , Lei Wang , Panrong Tong , Long Chen , Tianyu Xia , Gongjie Zhang , Xu Zhang , Chen Liu , Jianan Zhang , Chenglin Cai , Zhengwen Zeng , Steven Hoi , Yue Wang View a PDF of the paper titled OpenGCUA: Toward a Fully Open Stack for Generalist Computer Use, by Hanzhang Zhou and 14 other authors

View PDF HTML (experimental)

Abstract: Computer use has reached a turning point: the focus is expanding from GUI-only interaction to systems that complete tasks on users' computers through visual and programmatic interfaces. We define a generalist computer-use agent (GCUA) as a system in which the model chooses how to observe, act, and organize its work across visual and programmatic interfaces. Yet frontier implementations, such as GPT-6 Astra's computer-use system, remain largely closed, limiting systematic study of their design. We introduce Open Generalist CUA (OpenGCUA), an ongoing project to develop and open the full stack for GCUAs with the community. This report presents the first release of OpenGCUA. Code execution is a key feature of our agent design: it lets intermediate results determine subsequent operations and enables models to apply their existing programming capabilities directly to computer interaction. We also propose harness-to-model decision transfer, moving selected observations and task-state decisions from fixed harness rules to the model. On OSWorld-v2, our full configuration improves task scores across all eight evaluated models, including GPT-6 Astra and Claude Opus 5, raising mean partial score by +19.6 points and reducing reported steps per task by an average of 53.0% relative to batched GUI-only execution. We release our implementation and prompts, evaluation trajectories from all eight models, and preliminary versions of a real-device harness and a cross-device SDK. The harness addresses real-device challenges not fully captured by benchmarks, while the SDK provides a common programming interface across computers, mobiles and browsers. Across four real macOS tasks under our real-device harness, selected Qwen3.8-Max runs use 19.4 times as many reported reasoning tokens as GPT-6 Astra, highlighting substantial differences between models in real-device execution efficiency.

Subjects:

Computer Vision and Pattern Recognition (cs.CV)

Cite as: arXiv:2610.22200 [cs.CV]

(or arXiv:2610.22200v1 [cs.CV] for this version)

<https://doi.org/10.48550/arXiv.2610.22200>

Focus to learn more

arXiv-issued DOI via DataCite (pending registration)

Submission history From: Hanzhang Zhou [view email] [v1] Thu, 15 Oct 2026 16:51:35 UTC (11,885 KB)

Full-text links: Access Paper:

View a PDF of the paper titled OpenGCUA: Toward a Fully Open Stack for Generalist Computer Use, by Hanzhang Zhou and 14 other authors View PDF HTML (experimental) TeX Source

[view license](#)

[Additional Features](#)

[Audio Summary](#)

Current browse context:

cs.CV

[< prev](#)

[| next >](#)

[new](#) | [recent](#) | [2026-10](#)

Change to browse by:

[cs](#)

[References & Citations](#)

[NASA ADS](#) [Google Scholar](#)

[Semantic Scholar](#)

[export BibTeX citation](#) [Loading...](#)

[BibTeX formatted citation](#)

[×](#)

[loading...](#)

Data provided by:

[Bookmark](#)

[Bibliographic Tools](#)

[Bibliographic and Citation Tools](#)

[Bibliographic Explorer Toggle](#)

[Bibliographic Explorer \(What is the Explorer? \)](#)

[Connected Papers Toggle](#)

[Connected Papers \(What is Connected Papers? \)](#)

[Litmaps Toggle](#)

[Litmaps \(What is Litmaps? \)](#)

[scite.ai Toggle](#)

[scite Smart Citations \(What are Smart Citations? \)](#)

[Code, Data, Media](#)

[Code, Data and Media Associated with this Article](#)

[alphaXiv Toggle](#)

[alphaXiv \(What is alphaXiv? \)](#)

[Links to Code Toggle](#)

[CatalyzeX Code Finder for Papers \(What is CatalyzeX? \)](#)

DagsHub Toggle

DagsHub (What is DagsHub?)

GotitPub Toggle

Gotit.pub (What is GotitPub?)

Huggingface Toggle

Hugging Face (What is Huggingface?)

ScienceCast Toggle

ScienceCast (What is ScienceCast?)

Demos

Demos

Replicate Toggle

Replicate (What is Replicate?)

Spaces Toggle

Hugging Face Spaces (What is Spaces?)

Spaces Toggle

XYZ.AI (What is XYZ.AI?)

Related Papers

Recommenders and Search Tools

Link to Influence Flower

Influence Flower (What are Influence Flowers?)

Core recommender toggle

CORE Recommender (What is CORE?)

Author

Venue

Institution

Topic

About arXivLabs

arXivLabs: experimental projects with community collaborators

arXivLabs is a framework that allows collaborators to develop and share new arXiv features directly on our website.

Both individuals and organizations that work with arXivLabs have embraced and accepted our values of openness, community, excellence, and user data privacy. arXiv is committed to these values and only works with partners that adhere to them.

Have an idea for a project that will add value for arXiv's community? Learn more about arXivLabs .

Which authors of this paper are endorsers? | Disable MathJax (What is MathJax?)