

Mingyu Ouyang

yyyangwhu@gmail.com

yyyang404.github.io

(+65) 88851437

Google Scholar

LinkedIn

Research Focus Areas

My research focuses on multimodal digital agents that can perceive, plan, and ground its actions in digital environments. I am particularly interested in GUI and game agents, and the fine-grained perception and visual and action grounding needed for reliable agent-environment interaction.

Education

National University of Singapore

Ph.D. in Computer Science

Supervisors: Prof. Hwee Tou Ng and Prof. Mike Zheng Shou

Singapore

Jan. 2025 – Present

National University of Singapore

M.Sc. in Data Science and Machine Learning

Singapore

Aug. 2023 – Dec. 2024

Wuhan University

B.Sc. in Geospatial Informatics and Digitalized Technology

Wuhan, China

Sep. 2019 – Jul. 2023

GPA: 87.54/100; **Ranking:** 3/52 (2021-22), 16/52 (2020-21) in the annual assessment

Publications

[arXiv '26] GameWorld: Towards Standardized and Verifiable Evaluation of Multimodal Game Agents

Mingyu Ouyang*, Siyuan Hu*, Kevin Qinghong Lin, Hwee Tou Ng, Mike Zheng Shou

- Standardizes multimodal game-agent evaluation across 34 browser games and 170 tasks with paused sandbox execution, dual agent interfaces, and deterministic state-verifiable scoring.

[CVPR '26] FocusUI: Efficient UI Grounding via Position-Preserving Visual Token Selection

Mingyu Ouyang, Kevin Qinghong Lin, Mike Zheng Shou, Hwee Tou Ng

- Addresses the redundancy problem in UI preception. Selects instruction-relevant UI tokens while preserving positional continuity, improving grounding accuracy and efficiency under aggressive token reduction.

[AAAI '26] SlideTailor: Personalized Presentation Slide Generation for Scientific Papers

Wenzheng Zeng*, Mingyu Ouyang*, Langyuan Cui*, Hwee Tou Ng

- Generates user-aligned editable slides from papers using example-based preference conditioning, visual templates, and a chain-of-speech planning mechanism.

[arXiv '24] The Dawn of GUI Agent: A Preliminary Case Study with Claude 3.5 Computer Use

Siyuan Hu, Mingyu Ouyang, Difei Gao, Mike Zheng Shou

- Case study of Claude 3.5 Computer Use on diverse desktop tasks, with an out-of-the-box framework and analysis of planning, action, and critic limitations.

[CVPR '24] AssistGUI: Task-Oriented PC Graphical User Interface Automation

Difei Gao, Lei Ji, Zechen Bai, Mingyu Ouyang, Peiran Li, Dongxing Mao, Qinchen Wu, Weichen Zhang, Peiyi Wang, Xiangwu Guo, Hengxu Wang, Luwei Zhou, Mike Zheng Shou

- Introduces a 100-task Windows GUI benchmark and a GUI agent framework for task decomposition, GUI parsing, action generation, and reflection.

[TIP '24] JPEG Quantized Coefficient Recovery via DCT Domain Spatial-Frequential Transformer

Mingyu Ouyang, Zhenzhong Chen

- Introduces DCTransformer to recover JPEG-quantized DCT coefficients with joint spatial-frequency modeling across varied compression quality factors.

Professional Experience

Uselt AI

Co-founder

- Led the GUI agent development and workflow orchestration, focusing on creating, editing, and deploying personal agents.

Singapore

Mar. 2025 - Oct. 2025

Sensetime Singapore Research

Algorithm Researcher Intern

- Developed efficient low-level deep learning models for RAW image processing, including noise synthesis and learned Image Signal Processing for camera/phone in-built modules.

Singapore

Dec. 2023 - Oct. 2025

Tencent

Beijing, China

Machine Learning Engineer Intern

Mar. 2022 - Oct. 2022

- Implemented GPS data processing using Hadoop and Spark. Developed and maintained an internal service that automatically executes daily tasks, handling over **10 billion callbacks per hour**.

Springturn Education Technology

Zhuhai, China

Co-founder, Software Developer

Nov. 2020 - Sep. 2021

- Co-founded the company and secured an Angel Investment exceeding **¥ 1 million** in the second year of business.

Research Experience

ShowLab, NUS

Aug. 2023 – Present

- Research focus: MLLM-based agents, including GUI, computer-use, games; Agent-environment interaction.

IIP Lab, Wuhan University

Oct. 2022 – Jul. 2023

- Research focus: Low-level vision tasks, including image super-resolution, denoising, JPEG artifact removal.

State Key Lab of Info. Engineering in Surveying, Mapping and Remote Sensing

Sep. 2021 – June. 2022

- Research focus: 3D scene perception, fusing point cloud and BEV images for road scenes understanding.

Open-Source Projects

[GitHub 1.9k stars] Computer-Use OOTB: [showlab/computer_use_ootb](#)

- Core developer and maintainer. Desktop GUI agent for Windows/macOS with API-based and local execution.

[GitHub 1.8k stars] ShowUI: [showlab/showui](#)

- Maintainer. Lightweight VLA model for GUI agents.

Miscellaneous

Wuhan University Football Team

- Served as captain for three years and led the team to win the championship at the 16th Hubei Province Games (2022).

Technical Skills

- **Programming Languages:** Python (Advanced), Java (Hadoop, Spark), TeX, HTML/CSS (Vibe).
- **AI/ML Frameworks:** PyTorch, Transformers, vLLM, OpenAI compatible APIs.
- **Computer Use Agent Infra:** On-device control, cloud services (AWS, Azure), and agent framework orchestration.
- **Agentic Coding Assistance:** Cursor, Claude Code, Codex (spent 3B tokens per month with).