



# Technical Report of MoGUI and MoCon

Zichen Zhu\*, Liangtai Sun\*, Danyang Zhang\*,  
Ziyuan Li, Guangpeng Li, Lu Chen<sup>†</sup>, Kai Yu<sup>†</sup>  
X-LANCE Lab, Department of Computer Science and Engineering  
MoE Key Lab of Artificial Intelligence, SJTU AI Institute  
Shanghai Jiao Tong University, Shanghai, China  
{JamesZhutheThird,slt19990817}@sjtu.edu.cn

February 29, 2024

## 1 Introduction

ALL YOU NEED IS GUI-BASED MOBILE LLM AGENTS AND MORE DATA!  
IT IS VERYYYYYYYYY IMPORTANT FOR IMPROVING THE QUALITY OF OUR DAILY LIVES! SO WE ANNOUNCE THAT WE WILL OPEN OUR HARD-WORKING CRAWLED GUI DATASET CALLED MoGUI  AND MoCon .

IF YOU DO NOT KNOW WHAT IS GUI-BASED LLM AGENTS, WE REFER TO AppAgent[1], UFO[2], CogAgent[3], Mind2Web[4], ScreenAgent[5], AITW[6] AND THE MOST IMPORTANTLY, OUR WORKS META-GUI[7] AND CAM-GUI[8]!!!! (PLEASE CITE US, PLEASE ~~~)

---

\*Equal contribution.

<sup>†</sup>Corresponding authors.

<sup>‡</sup>THIS IS NOT THE OFFICIAL RELEASE.

## References

- [1] Zhao Yang, Jiaxuan Liu, Yucheng Han, Xin Chen, Zebiao Huang, Bin Fu, and Gang Yu. Appagent: Multimodal agents as smartphone users. *arXiv preprint arXiv:2312.13771*, 2023.
- [2] Chaoyun Zhang, Liqun Li, Shilin He, Xu Zhang, Bo Qiao, Si Qin, Minghua Ma, Yu Kang, Qingwei Lin, Saravan Rajmohan, et al. Ufo: A ui-focused agent for windows os interaction. *arXiv preprint arXiv:2402.07939*, 2024.
- [3] Wenyi Hong, Weihang Wang, Qingsong Lv, Jiazheng Xu, Wenmeng Yu, Junhui Ji, Yan Wang, Zihan Wang, Yuxiao Dong, Ming Ding, et al. Cogagent: A visual language model for gui agents. *arXiv preprint arXiv:2312.08914*, 2023.
- [4] Xiang Deng, Yu Gu, Boyuan Zheng, Shijie Chen, Sam Stevens, Boshi Wang, Huan Sun, and Yu Su. Mind2web: Towards a generalist agent for the web. *Advances in Neural Information Processing Systems*, 36, 2024.
- [5] Runliang Niu, Jindong Li, Shiqi Wang, Yali Fu, Xiyu Hu, Xueyuan Leng, He Kong, Yi Chang, and Qi Wang. Screenagent: A vision language model-driven computer control agent. *arXiv preprint arXiv:2402.07945*, 2024.
- [6] Christopher Rawles, Alice Li, Daniel Rodriguez, Oriana Riva, and Timothy Lillicrap. Android in the wild: A large-scale dataset for android device control. *arXiv preprint arXiv:2307.10088*, 2023.
- [7] Liangtai Sun, Xingyu Chen, Lu Chen, Tianle Dai, Zichen Zhu, and Kai Yu. Meta-gui: Towards multi-modal conversational agents on mobile gui. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 6699–6712, 2022.
- [8] Zichen Zhu, Liangtai Sun, Jingkai Yang, Yifan Peng, Weilin Zou, Ziyuan Li, Wutao Li, Lu Chen, Yingzi Ma, Danyang Zhang, et al. Cam-gui: A conversational assistant on mobile gui. In *National Conference on Man-Machine Speech Communication*, pages 302–315. Springer, 2023.