





Building a GUI data annotation tool for BIM authoring software

Modeling:	
Mathematics:	
Programming:	
Science:	

Description

This project aims to develop a cross-platform (Windows + macOS) GUI recording and annotation tool to collect high-quality human demonstrations in BIM authoring software (e.g., Vectorworks), and then use the collected dataset to train a simple computer-use / behavior cloning model.

Recent benchmarks [3] of computer-use agents and data-collection frameworks show that robust training data should align screen visuals, user actions, and UI structure. OSWorld [1] highlights evaluation settings across Windows/macOS/Linux and stresses GUI grounding challenges. OpenCUA [2] further provides an open foundation for data collection and uses a cross-platform recorder (AgentNetTool) that captures synchronized screen video, mouse/keyboard events, and accessibility trees, and explicitly mentions using OBS Studio for screen recording.

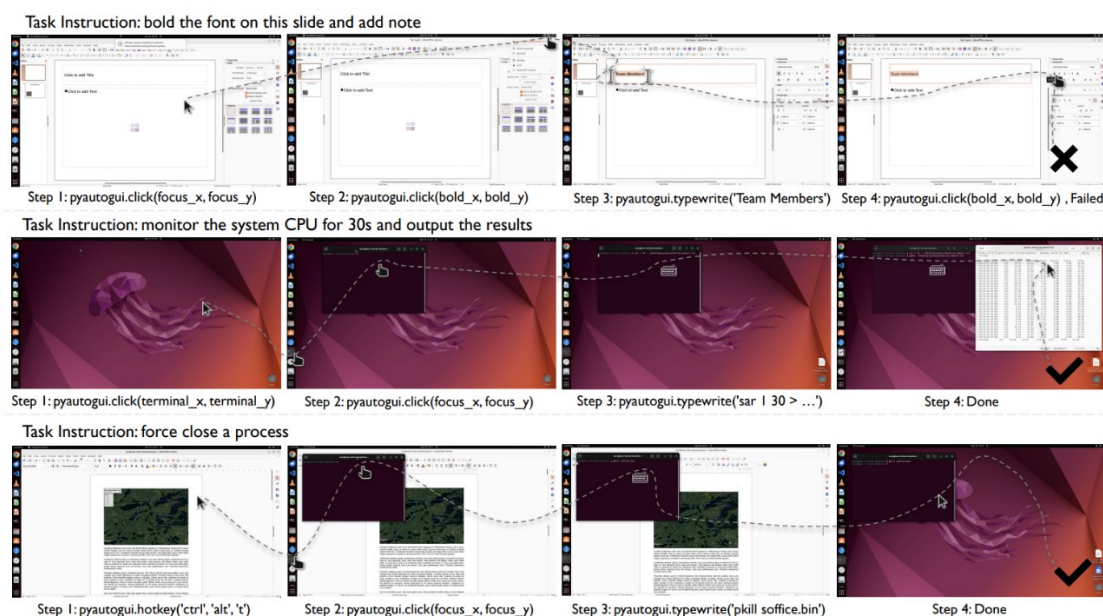


Figure 1 Example GUI trajectory in OSworld [1]

Task

- Literature and tooling review on computer-use model, datasets and capture stacks, focusing on how they represent state/action and UI grounding.
- Design a dataset schema and implement the annotation tool.
- Build a pilot dataset that may include: mouse and keyboard actions, video recordings, screenshots with timestamps, UI button/widget information, reasoning traces, software command logs, etc.
- Train the model using the collected data.
- Evaluate both the data annotation tool and the trained model.

Supervisors

Changyu Du and Zihan Deng / Chair of Computing in Civil and Building Engineering
changyu.du@tum.de, zihan.dengr@tum.de

References

- [1] Xie, T., et al. OSWorld: Benchmarking Multimodal Agents for Open-Ended Tasks in Real Computer Environments. NeurIPS 2024.
- [2] Wang, X., et al. OpenCUA: Open Foundations for Computer-Use Agents.
- [3] Man, Brandon, et al. "VideoCAD: A Large-Scale Video Dataset for Learning UI Interactions and 3D Reasoning from CAD Software." arXiv preprint arXiv:2505.24838 (2025).