An Awesome Desktop Environment

Author 1 Author 2 Author 3

**Abstract** In this report, we are honored to introduce a novel awesome interaction environment for desktop agents --- OSWorld. This environment is build upon virtual machine technology to provide the most adaptable infrastructure that can port to any architectures of host machine. We define basic atomic actions and offer screenshots and accessibility trees (AT) as observations, which constitute a universal interfaces for the most diverse agents. A lot of efforts are dedicated to collecting and crafting diverse and realistic computer tasks. These tasks raise from in-the-wild need of human posted to real-world web communities and reflect real characteristics of everyday computer use. The execution result of each tasks is assessed accurately by a hand-crafted evaluation script. We hope this new environment will bring a new trend for the research community of GUI and decision-making.

# 1. Introduction

TODO

# 2. Related Work

**Benchmarks for Vision-Language Digital Agents** TODO

**Vision-Language Models for Digital Agents** TODO

# 3. Environment Structure & Features

## 3.1 Internal Structure

TODO

## 3.2 Observation Space & Action Space

TODO

# 4. Open-Ended Benchmark Suite

## 4.1 Task Definition

TODO

## 4.2 Data Collection

TODO

## 4.3 Initial State Setup

TODO

## 4.4 Example-Specific Evaluation

TODO

# 5. Benchmarking Advanced Models and Methods

## 5.1 Methods

TODO

## 5.2 Main Results

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| model | os | calc | impress | writer | vlc | gimp | chrome | thunderbird | vscode | multi | avg |
| Gpt-4 | 0.3333 | 0.0000 | 0.0910 | 0.0476 | 0.1507 | 0.0625 | 0.3636 | 0.1333 | 0.4545 | 0.2143 | 0.1557 |

## 5.3 Qualitative Analysis

TODO

# 6. Conclusion

TODO