

---

# Supplementary Materials

---

## 1 Documentation and Intended Uses

This dataset covers six GUI scenarios and eight types of GUI-orientated questions in three formats, with 12,379 videos and more than 100k QA pairs focusing on both static and dynamic GUI content. All features are listed and explained below:

- video: the video of the GUI content.
- question: each video has 6 questions, with 4 free-form questions and 2 multiple-choice questions.
- golden answer: each free-form question has a golden answer, while the multiple-choice question has a definite correct option.
- multi-round conversation: each video has a multi-round conversation, with two roles: 'Assistant' and 'User'.
- human-selected keyframe: each video contains human-selected keyframes (except android), with annotation of subgoal, operation of mouse and keyboard.
- description (caption) of video: each video has two detailed captions and one concise caption, focusing on different aspects of the video content.
- description of keyframes (only for test split): In test split, we use GPT-4V to annotate all the keyframe for detailed and concise caption for ablation study.

This dataset will be used for benchmarking current MLLMs, pre-training datasets for GUI-Orientated models, and future research in various GUI-related areas, such as GUI grounding and General Virtual Agents.

## 2 Accessibility

All the relevant URLs are listed below:

	URL
Home Page	<a href="https://gui-world.github.io/">https://gui-world.github.io/</a>
DataSet	<a href="https://huggingface.co/datasets/shuaishuaicdp/GUI-World">https://huggingface.co/datasets/shuaishuaicdp/GUI-World</a>
Github Codebase	<a href="https://github.com/Dongping-Chen/GUI-World">https://github.com/Dongping-Chen/GUI-World</a>

Our datasets are publicly available on Hugging Face. You can view the data items directly on the webpage. The metadata and specifications of the datasets are presented in the datacard. You can conveniently download the dataset using the following Python code:

```
29 # Please install the datasets package first by using the command 'pip install datasets'.
30 from datasets import load_dataset
31 # Loading Dataset
32 dataset = load_dataset("shuaishuaicdp/GUI-World")
```

### 33 **3 Reproducibility of the Benchmark**

34 To replicate the evaluation results presented in our paper, please follow these steps  
35 (the detailed instructions can be found in the Readme.md in the supplementary  
36 materials):

- 37 1. Download our GUI-Vid model weights following instructions in the GitHub  
38 repository: GUI-World Github.
- 39 2. Obtain our GUI-World dataset from the following link: GUI-World Dataset.
- 40 3. Visit and clone our GitHub repository at Github Codebase.

### 41 **4 Author Statement**

42 As the authors of this dataset, we hereby declare that we bear full  
43 responsibility for any violations of rights, including but not limited  
44 to intellectual property rights, privacy rights, or any other legal  
45 rights that may arise from the distribution of this dataset. We  
46 assure that the dataset complies with all ethical guidelines and  
47 legal requirements in the creation and distribution of the dataset.  
48 We confirm that the dataset is distributed under the [CC BY 4.0]  
49 license.

### 50 **5 Plan**

51 The dataset and benchmark model weights have been uploaded to HuggingFace, and  
52 the dataset has been publicly released under the [CC-BY-4.0] license. HuggingFace  
53 is designated as the primary release platform, with plans for ongoing optimization  
54 and expansion of the dataset. Prior to the paper's official publication, the dataset  
55 will be archived, and a Digital Object Identifier (DOI) will be assigned. Once a DOI is  
56 assigned to the dataset on HuggingFace, it becomes immutable, preventing actions  
57 such as deletion, renaming, transfer, or modification of visibility to private. This  
58 mechanism ensures the dataset's enduring preservation and affords it a persistent,  
59 dereferenceable identifier.

60 **Note: We have decided to change the license from MIT mentioned in the paper**  
61 **to CC-BY-4.0**