

Responsible NLP Checklist

Paper title: *Towards Scalable Lightweight GUI Agents via Multi-role Orchestration*

Authors: *Ziwei Wang, Junjie Zheng, Leyang Yang, Sheng Zhou, Xiaoxuan Tang, Fang Zhouhua, Zhiwei Liu, Dajun Chen, Yong Li, Jiajun Bu*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- the authors indicated that the question does not apply to their work
- the authors did not respond to the checkbox question

For background on the checklist and guidance provided to the authors, see the [Responsible NLP Checklist](#) page at ACL Rolling Review.

A. Questions mandatory for all submissions.

- A1. Did you describe the limitations of your work?

This paper has a Limitations section.

- A2. Did you discuss any potential risks of your work?

This work is a scientific exploration; the data and testing environment used are publicly available and reliable, and there are no potential risks.

B. Did you use or create scientific artifacts? (e.g. code, datasets, models)

- B4. Did you discuss the steps taken to check whether the data that was collected/used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect/anonymize it?

This work is a scientific exploration; the data and testing environment used are publicly available and reliable. It does not contain data containing personally identifiable information or offensive content.

- B6. Did you report relevant statistics like the number of examples, details of train/test/dev splits, etc. for the data that you used/created?

A.1 Data Curation & A.6 Prompt Templates & A.7 Case Study

C. Did you run computational experiments?

- C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?

5.1 Implementation Details

- C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?

5. Experiments & A.6 Prompt Templates & A.7 Case Study

The Responsible NLP Checklist used at ACL Rolling Review is adopted from NAACL 2022, with the addition of ACL 2023 question on AI writing assistance and further refinements based on ARR practice. ACL 2026 used a subset of ARR checklist form.

D. Did you use human annotators (e.g., crowdworkers) or research with human subjects?

- D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?

In this work, we did not employ any human staff for data annotation; all data synthesis involved utilized open-source data and models for post-processing, and the instructions are detailed in "A.6 Prompt Templates".

- D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?

In this work, we did not employ any human staff for data annotation; all data synthesis involved utilized open-source data and models for post-processing, and the instructions are detailed in "A.6 Prompt Templates".

- D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating (e.g., did your instructions explain how the data would be used)?

A.1 Data Curation & A.2 Benchmark Details and Metrics. The data and models used in this paper are all open source models. We strictly adhere to the restrictions stipulated in the license agreement and provide citations of the artifacts in the paper.

- D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?

A.1 Data Curation & A.2 Benchmark Details and Metrics & A.6 Prompt Templates. The data and models used in this paper are all open-source models. We strictly adhere to the restrictions stipulated in the license agreement and cite all sources in this paper. We also explain the data collection strategy and data synthesis instructions in "A.1 Data Curation" and "A.6 Prompt Templates".

E. Did you use AI assistants (e.g., ChatGPT, Copilot) in your research, coding, or writing?

- E1. If you used AI assistants, did you include information about their use?

We did not use AI assistants to help us with this work (coding and writing). Our data synthesis strategy used publicly available MLLMs, and we described the instruction details in "A.6 Prompt Templates".