

Responsible NLP Checklist

Paper title: *DaMo: Data Mixing Optimizer in Fine-tuning Multimodal LLMs for Mobile Phone Agents*
Authors: *Kai Shi, JunYang, Ni Yang, Binqiang Pan, Qingsong Xie, Zhangchao, Zhenyu Yang, Tianhuang Su, Haonan Lu*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- N/A 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 focuses on algorithmic optimization of data mixture for multimodal LLMs in mobile agent scenarios, which does not involve direct ethical risks, privacy issues, or potential harms to individuals/society, so no dedicated discussion of potential risks is included.

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?

All datasets used in this work (including the self-constructed PhoneAgentBench and open-source training datasets) are manually curated and verified, containing no personally identifiable information (PII), offensive content, or sensitive data, so no additional anonymization steps are required.

- 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?

We report detailed statistics for all datasets used in this work, including the self-constructed PhoneAgentBench (1235 samples, 6 task-specific datasets) and open-source training datasets, in Appendix B and Section 3.

C. Did you run computational experiments?

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

All experimental results are reported as the average of multiple independent runs, with detailed descriptive statistics and transparency on the reported metrics (mean values) provided in Section 5.

- 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?

All experimental results are reported as the average of multiple independent runs, with detailed descriptive statistics and transparency on the reported metrics (mean values) provided in Section 5.

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.?

We provide the detailed annotation guidelines and instructions given to internal annotators for constructing the PhoneAgentBench dataset in Appendix A.

- 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)?

All annotators are full-time employees of OPPO, recruited through internal project assignment, and compensated in accordance with OPPO's standard employee compensation policies, which are adequate for their demographic and regional context.

- 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)?

All annotators provided informed consent for their annotation work, and were fully informed that the data would be used for academic research and publication, with all personal information anonymized in the final dataset.

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

This research involves internal employees of OPPO for data annotation, conducted in accordance with OPPO's internal research ethics and data protection policies. The work does not involve external human subjects research requiring formal IRB/ethics board approval, and all annotators provided informed consent for their participation.

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 disclose the use of AI assistants in the Appendix E of the paper.