

## Responsible NLP Checklist

Paper title: *Test-Time Scaling in Multimodal Foundation Models: A Comprehensive Survey of Generation and Reasoning*

Authors: *Cong Wan, Ying He, Zhongzhan Huang, Hefeng Wu*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- <sup>N/A</sup> 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 paper is a survey of existing literature and does not introduce new models, datasets, or experimental systems. Therefore, it does not directly create additional risks beyond those already present in the surveyed works.*

### 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 literature survey and does not involve new data collection or direct use of raw data. Therefore, we did not conduct an additional analysis of personally identifying information or offensive content beyond what is reported in the original sources.*

- 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?

*This work is a literature survey rather than an empirical dataset paper. We do not create or use a single dataset under a unified experimental setup, so statistics such as the number of examples and train/dev/test splits are not reported in a standardized form.*

### C. Did you run computational experiments?

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

*This paper is a survey of existing literature and does not create new datasets or use a single dataset under a unified experimental setup. Therefore, we do not report dataset statistics such as the number of examples or train/dev/test splits in the same way as empirical papers.*

- 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?

---

*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.*

*This paper is a survey of existing literature and does not report original experimental results from repeated runs under a unified setup. Therefore, we do not provide descriptive statistics such as error bars, summary statistics across runs, or distinctions between max/mean results in the manner expected for empirical studies.*

**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.?

*This paper is a literature survey and does not involve human participants, annotators, or user studies. Therefore, there were no participant instructions, screenshots, or risk disclosures to report.*

- 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)?

*This paper is a literature survey and does not involve recruiting or compensating participants or annotators. Therefore, no recruitment or payment information is applicable.*

- 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)?

*This paper is a literature survey and does not involve new data collection, curation, or direct use of participant-level data by the authors. Therefore, we do not discuss consent procedures beyond what is reported in the original sources.*

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

*This paper is a literature survey and does not involve new data collection, human subjects research, or annotation studies conducted by the authors. Therefore, no ethics review board approval was required for this work.*

**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?

*AI assistants were used only for limited writing and editing support. All technical content, judgments, and final revisions were performed and verified by the authors. Because the use was limited and did not contribute original scientific content, we did not include a separate statement in the paper.*