

Responsible NLP Checklist

Paper title: *Rhombus: Incentivizing Coordination in Parallel Thinking through Reinforcement Learning*
Authors: *Ziyuan Nan, Qi Yi, Di Huang, Yutong Wu, Guanhua Huang, Xue Gong, Kejiao Li, Yuhao Jiang, Chenchen Zhang, Zenan Xu, Xing Hu, Bo Zhou*

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?

We primarily focus on improving mathematical and algorithmic reasoning using established, open-source benchmarks. Our method optimizes the efficiency and accuracy of existing Large Language Models without introducing new datasets, identifying information, or training objectives that would encourage the generation of harmful, biased content.

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?

Our work utilizes established, publicly available benchmarks focused exclusively on mathematical reasoning and algorithmic coding. These datasets consist of technical problem statements and solutions; by their nature, they do not contain 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?

Section 4.1

C. Did you run computational experiments?

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

Section 4.4

- 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?

Section 4.1, Appendix A

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.?

(left blank)

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)?

(left blank)

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)?

(left blank)

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

(left blank)

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 utilized AI assistance exclusively for linguistic refinement of the authors' original content and for facilitating literature search.