

Responsible NLP Checklist

Paper title: *Chain-of-Thought as a Lens: Evaluating Structured Reasoning Alignment between Human Preferences and Large Language Models*

Authors: *Boxuan Wang, Zhuoyun Li, Xinmiao Huang, Xiaowei Huang, Yi Dong*

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?

Ethical Considerations section

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?

Ethical Considerations section

- 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 4.1 and Appendix A.2

- 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

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 did not recruit external participants, annotators, or crowd-workers. The only annotation-related human involvement was internal author-side sanity checking and quality assurance of reference chains constructed from public QA benchmarks and AI-assisted generation; see Appendix D. Therefore, there were no participant instructions to report.

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.

N/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)?

No external participants or annotators were recruited or paid. The manual checking was performed internally by the authors for quality assurance.

N/A 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)?

We did not collect personal data from individuals or curate a human-subject dataset. The work uses publicly available benchmark items and author-side sanity checks, so participant consent was not applicable.

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

No external human-subject study, participant recruitment, or personal data collection was conducted. Any annotation-related human involvement was limited to internal author-side sanity checks and quality assurance of public benchmark items.

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 used AI assistants to support language polishing (e.g., improving clarity and grammar). We did not specifically include this information because: all technical ideas, methods, experiments, analyses, and conclusions in this paper are original work by the authors. Any external materials were independently verified and are properly cited.