

Responsible NLP Checklist

Paper title: *Triviality Corrected Endogenous Reward*

Authors: *Xinda Wang, Zhengxu Hou, Yangshijie Zhang, yanbingren, Jialin Liu, ChenZhuo Zhao, Zhibo Yang, Bin-Bin Yang, Feng Xiao*

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?

Appendix C.3. In the interpretability section where we score reward values, we use multiple proprietary models in aggregation to evaluate highlighted sentences, which carries some risk of bias.

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

- ^{N/A} 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?

N/A. We use the publicly available DeepWriting and LongWriter datasets, which do not contain personal 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. We report the training data splits and test benchmarks in the experimental section, with detailed descriptions provided in the appendix.

C. Did you run computational experiments?

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

Appendix B.3. We report the specific values of parameters and , along with complete hyperparameter tuning experiments.

- 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. In the experimental section, we report that for writing tasks, we average scores from three LLM judge evaluations. For mathematical tasks, we use different metrics across benchmarks: for AIME 2024, AIME 2025, and AMC, which have smaller test sets, we report the avg@32 metric; for the other three larger benchmarks, we report standard pass@1 accuracy.

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.?

N/A. We did not use human annotators (e.g., crowdworkers) or conduct experiments with human subjects.

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)?

N/A. We did not use human annotators (e.g., crowdworkers) or conduct experiments with human subjects.

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)?

N/A. We did not use human annotators (e.g., crowdworkers) or conduct experiments with human subjects.

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

N/A. We did not use human annotators (e.g., crowdworkers) or conduct experiments with human subjects.

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?

(left blank)