

Responsible NLP Checklist

Paper title: *Reasoning-Guided Exploration for Online DPO*

Authors: *Zetian Hu, Shunyu Liu, Ting-En Lin, Fei Huang, Yongbin Li, Dacheng Tao*

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?

Section 6 (Limitations)

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 uses publicly available, pre-existing benchmark datasets that have already undergone data curation and processing by their original creators.

- 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?

The paper utilizes existing publicly available datasets (RLHFlow, HH-RLHF, SHP, HelpSteer, Ultra-Feedback, CodeUltraFeedback, UltraInteract) and evaluation benchmarks (AlpacaEval 2.0, ARC-Challenge, GSM8K, etc.) that have already reported relevant statistics in their original publications.

C. Did you run computational experiments?

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

We did not perform hyperparameter search in this work. Following standard practices in online DPO research, we adopted commonly used hyperparameter settings from prior work to ensure fair comparison with baseline methods. The focus of our contribution is the novel dual-reward evaluation mechanism and weighted ranking preference construction method, not hyperparameter optimization. Conducting extensive hyperparameter search would have required substantial additional computational resources and would have made it difficult to isolate whether performance improvements stem from our proposed method or from better-tuned hyperparameters.

- 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

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?
AI tools were used only for supplementary tasks (e.g., polishing language, checking grammar, generating boilerplate code structures) that did not contribute to the intellectual content or findings of the research.