Responsible NLP Checklist

Paper title: *Topology-Enhanced Alignment for Large Language Models: Trajectory Topology Loss and Topological Preference Optimization*
Authors: *Yurui Pan, Ke Xu, Bo Peng*

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?
*Yes. We discuss potential risks of alignment objectives and bias amplification in the **Ethics Statement** section, and also touch on remaining safety concerns in the **Limitations** 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?
*No. We did not run additional checks beyond those performed by the original dataset providers (UltraChat and HH-RLHF), which already released the data as public benchmarks after their own filtering and anonymization. We note this limitation in the **Ethics Statement** and **Limitations** sections.*

☑ 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?
*Yes. We describe the datasets used (UltraChat for SFT and HH-RLHF for DPO) and their splits in **Section 4.1**, including the number of training and evaluation examples for each setting.*

☑ C. Did you run computational experiments?

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
*Yes. We describe the experimental setup, model, and training configurations in **Section 4.1**, and provide additional hyperparameter details (learning rates, batch sizes, LoRA settings, topology-related coefficients, etc.) in **Appendix H***

☑ 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?
in section 4

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?**

N/A 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 or pay any new human participants; we only use publicly available datasets (UltraChat and HH-RLHF) that were collected and released by other organizations.

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)?
We did not recruit or pay any new human participants; we only use publicly available datasets (UltraChat and HH-RLHF) that were collected and released by other organizations.

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 recruit or pay any new human participants; we only use publicly available datasets (UltraChat and HH-RLHF) that were collected and released by other organizations.

N/A D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
We did not recruit or pay any new human participants; we only use publicly available datasets (UltraChat and HH-RLHF) that were collected and released by other organizations.

☒ **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 did not use generative AI in this paper.