

## Responsible NLP Checklist

Paper title: *ClinAlign: Scaling Healthcare Alignment from Clinician Preference*

Authors: *Shiwei Lyu, Xidong Wang, Hao Zhu, Lei Liu, Chaohe Zhang, Jian Wang, Jinjie Gu, Benyou Wang, Yue Shen*

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?

*Although the paper discusses safety-oriented alignment strategies and some methodological limitations, it does not explicitly provide a dedicated discussion of the broader risks of deploying this work. Such risks include incorrect or overconfident medical advice, over-reliance on model outputs in high-stakes settings, and the possibility that improved alignment may still not eliminate failures on reasoning-intensive cases.*

### 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?

*The paper does not explicitly discuss checks for personally identifying or offensive content. The data used are publicly available open-source datasets and do not contain personally identifying information.*

#### 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?

*Sections 3, 4.1, 4.2, and 5.1.*

### C. Did you run computational experiments?

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

*Section 5.1*

#### 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?

*The task involves large-model experiments with substantial computational cost, which limited our ability to perform multiple repeated runs. Therefore, we report single-run results rather than descriptive statistics over multiple trials.*

---

*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.?

*The paper reports compensation information, but it does not describe the recruitment process in detail or discuss whether the compensation was adequate given the participants demographic background.*

- 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)?

*The annotation work was carried out by internal employees. We do not provide detailed information about recruitment procedures or compensation, as these are internal operational details subject to commercial confidentiality.*

- 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)?

*The paper does not explicitly discuss whether or how consent was obtained from people whose data were included in the source datasets.*

- 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?

*Sections 3, 4, and Appendices AF.*