Responsible NLP Checklist

Paper title: *FaithLens: Detecting and Explaining Faithfulness Hallucination*
Authors: *Shuzheng Si, Qingyi Wang, Haozhe Zhao, Yuzhuo Bai, Guanqiao Chen, Kangyang Luo, Gang Chen, Fanchao Qi, Minjia Zhang, Baobao Chang, Maosong Sun*

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?
Our work is designed to build hallucination detection models, thus there is no potential risks for research communities.

☑ **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?
Our collected data is from open-source communities that does not contain any information that names or uniquely identifies individual people.

☑ 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?
We provide the detailed introduction of our used datasets in the Appendix and Section Experiments.

☑ **C. Did you run computational experiments?**

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
We provide the detailed introduction of experimental setup and hyperparameters in the Appendix and Section 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?
We provide the detailed introduction of descriptive statistics in the Appendix and Section Experiments.

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.?
We provide the full text of instructions in the Appendix.

☑ 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 provide the such information in the Appendix.

☑ 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 provide the such information in the Appendix. Meanwhile, we will open-source our training data.

☑ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
Our collected data is from open-source communities, thus there is no ethics issue.

☑ **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?
Privately revealed to ACL ARR 2026 January Program Chairs, ACL ARR 2026 January Submission1149 Area Chairs, ACL ARR 2026 January Submission1149 Authors, ACL ARR 2026 January Submission1149 Reviewers, ACL ARR 2026 January Submission1149 Senior Area Chairs We use AI assistants in our experiments for comparisons, e.g., GPT-4.1. We provide the related information in the Appendix and Section Experiment.