

## Responsible NLP Checklist

Paper title: *Verifiable LLM-Generated Text Detection via Projected Semantic-Structural Distributions*  
Authors: *Ruochong Xiong, Qien Li, Wangwang Lian, Yulong Wan, Hanlin Xue, Zhouxing Tan, Han Yang, Fengyu Lu, Junfei Liu*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- <sup>N/A</sup> 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.*

<sup>N/A</sup> A2. Did you discuss any potential risks of your work?  
*(left blank)*

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

<sup>N/A</sup> 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?  
*(left blank)*

<sup>N/A</sup> 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?  
*An overview of the benchmark is presented in Section 4.1. Descriptive statistics and detailed descriptions of the benchmark dataset are provided in Appendices D.1 and D.2, while data regarding the supplementary evaluation set can be found in Appendix E.6.*

### C. Did you run computational experiments?

C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?  
*The overall experimental setup is introduced in Section 4.1. Details regarding baseline parameter settings are provided in Appendix D.3, the complete experimental settings for our method are detailed in Appendix F. Section 4.3 and 4.4 presents the hyperparameter analysis and ablation studies.*

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 reporting details for the main experiments are provided in Section 4.1 and Appendix D.1. Additionally, the settings for multiple repeated experiments, conducted to demonstrate statistical validity, are detailed in section 4.4, Appendices E.4 and F.*

*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 only use annotators to spot-check whether the data quality meets requirements. Specifically, the entire inspection procedure is conducted in strict accordance with the protocols outlined in the original DetectRL paper.*

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?

*We employed large language models solely as general-purpose writing assistants to refine the manuscript. Specifically, LLMs were used to correct grammar, enhance clarity, and improve phrasing in certain sentences. The models did not contribute to the research design, problem formulation, method development, experimentation, analysis, or overall scientific contributions. Their role was strictly limited to surface-level editing and presentation improvements of the paper.*