

## Responsible NLP Checklist

Paper title: *Look Twice before You Leap: A Rational Framework for Localized Adversarial Text Anonymization*

Authors: *Donghang Duan, Xu Zheng, Yuefeng He, Chong Mu, Leyi Cai, Lizong Zhang*

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?

*Ethics Statement (Page 10). We discuss the potential dual-use risks of LLMs and clarify that our framework is designed strictly for defensive evaluation and enhancing user privacy through localized deployment.*

### 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?

*Section 4.1 (Page 5, 6), Ethics Statement (Page 10). We exclusively use established public datasets and strictly adhere to ethical standards by avoiding any new collection of private data.*

- 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?

*Section 4.1 (Page 5, 6). We report the details of the two datasets used.*

### C. Did you run computational experiments?

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

*Section 4.1 (Page 5, 6), Appendix C (Page 13, 14, 15). We detail the hardware environment, local model backbones, while Table 7 provides a threshold selection analysis, Table 8 provides comprehensive training hyperparameters and Table 9 provides generation hyperparameters to ensure complete reproducibility.*

- 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 4.3 (Page 6, 7, 8, 9). We report mean scores for all quantitative metrics across multiple runs in Table 1 and 2. For human evaluation, we provide the average win rate along with standard deviation in Figure 5 to reflect inter-annotator consistency.*

---

*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.?

*Section 4.3.4 (Page 9). We detail the human evaluation protocol, which involves a blind three-way pairwise comparison by five independent annotators and a human validation study with three annotators with NLP research backgrounds.*

- 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)?

*Section 4.3.4 (Page 9). We recruited a total of eight independent, non-author volunteer annotators from the research community. As the evaluation was a voluntary scholarly contribution within the research community, no financial compensation was provided.*

- 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)?

*Ethics Statement (Page 10). We rely exclusively on established public datasets and strictly adhere to ethical standards by avoiding any new collection of private data.*

- 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?

(left blank)