

Responsible NLP Checklist

Paper title: *Breaking the Generator Barrier: Disentangled Representation for Generalizable AI-Text Detection*

Authors: *Xiao Pu, Zepeng Cheng, Lin Yuan, Yu Wu, Xiuli Bi*

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.

- N/A* A2. Did you discuss any potential risks of your work?

N/A

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

- N/A* 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?

N/A

- 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 reported dataset statistics in Section 4.1 (Datasets) and Table 1, including the number of AI-generated and human-written samples across 20 LLMs and 7 categories, as well as the balanced sampling strategy. Details of train/test splits under the leave-one-generator-out setting are also described in Section 4.2.1, and also in appendix.

C. Did you run computational experiments?

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

We discussed the experimental setup in Section 4.2 (Experimental Setup), including optimizer choice (AdamW), batch size (16), learning rate (210), and sampling strategy. We reported the selected hyperparameter values but did not conduct an extensive hyperparameter search.

- 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 reported descriptive statistics in Section 5 (Results) and Tables 25, including accuracy and F1 scores under the leave-one-generator-out setting. The reported results are from single runs and are directly comparable across methods.

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.?

This work does not involve human subjects or annotators. We only used existing benchmark datasets (MAGE), hence no participant instructions apply.

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)?

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)?

N/A

D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?

N/A

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 used AI assistants (e.g., ChatGPT) for language polishing and grammar refinement in writing. All ideas, analyses, experiments, and claims are the responsibility of the authors.