

Responsible NLP Checklist

Paper title: *TAMA: Target-Aware Multilingual Abuse Detection by Cascaded Conditional Multi-Task Learning*

Authors: *JiYan Liu, Youzheng Liu, Taihang Wang, Yimin Wang, Ye Jiang, Diana Maynard*

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?

We include a content warning/disclaimer that the examples in this paper may be offensive or vulgar (Disclaimer on page 1), and we discuss the broader harm context of coordinated harassment against public figures in the Introduction (Section 1).

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?

We explicitly warn that the paper contains offensive or vulgar examples (Abstract). The data is collected for targeted abuse detection against public figures (Section 3.13.2), and we perform cleaning to remove deleted/unavailable posts and duplicates before annotation (Section 3.2).

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 report dataset statistics (e.g., number of instances and label distribution) and the train/val/test split in Section 3 (Dataset) and Appendix A (Experimental Setup).

C. Did you run computational experiments?

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

Appendix A (Experimental Setup).

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 report single-run test results using the best development checkpoint, and do not include error bars or statistics across multiple random seeds/runs.

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.?

3. Dataset, Appendix C and Appendix D

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 describe recruiting four volunteers and note that annotation was uncompensated and voluntary in Sec. 3.3

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 did not obtain individual consent from post authors; the data are derived from publicly available posts. We mitigate risks by data minimization, platform-compliant release practices (e.g., sharing identifiers), and honoring deletions/unavailable posts.

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

We did not seek ethics approval because the dataset is derived from publicly available posts and involves no interaction or intervention with individuals. We mitigate risks via data minimization, avoiding unnecessary reproduction of offensive content, platform-compliant release where applicable, and removing deleted/unavailable posts.

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 an LLM (GPT-5) to assist annotation by proposing candidate evidence spans, while all final labels/spans were verified and finalized by human annotators. We also used a safety classifier (Granite Guardian) for pre-filtering during data collection; these tools were used as assistance only and did not serve as ground-truth annotations (see Sec. 3.23.3).