

Responsible NLP Checklist

Paper title: *Scaling Laws or Threshold Effects: Exploring the Optimal Vocabulary Size for Balancing Performance and Efficiency in Low-Resource Languages*

Authors: *Ao Han, Andong Chen, Yuan Sun, Xiaobing Zhao*

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?

We discussed the ethical considerations and potential risks in Section 8 (Ethics Statement). Specifically, we addressed linguistic justice and the mitigation of tokenization disparity for minority languages.

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 discussed data safety and privacy in Section 8 (Ethics Statement). The datasets used (MC2) are publicly available and we adhered to their original licenses and privacy protocols, ensuring no sensitive personal data was involved.

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?

Relevant statistics including vocabulary expansion scales and language-specific token counts are reported in Table 1 (Section 3.1) and Table 5 (Appendix A.1). Dataset benchmarks are detailed in Section 3.3.

C. Did you run computational experiments?

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

The experimental setup, including training protocols, model architectures, and hardware platforms, is discussed in Section 3.3. Detailed hyperparameters for pre-training (PT) and fine-tuning (SFT) are provided in Table 6 and Appendix A.3.

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,

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.

etc. or just a single run?

We reported average values and performance trajectories across multiple tasks. As specified in Appendix A.4, each experiment was repeated three times, and we reported the average values of metrics. Statistical significance (p-values) is discussed in Section 4.1 (The BBPE Threshold Effect).

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 study does not involve human subjects or manual annotation.

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)?

This study does not involve human subjects or manual annotation.

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)?

This study does not involve human subjects or manual annotation.

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

This study does not involve human subjects or manual annotation.

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?

AI assistants were used to polish the linguistic quality and clarity of the manuscript. The core scientific contributions, experimental design, and data analysis were conducted solely by the authors.