

## Responsible NLP Checklist

Paper title: *SPEAK: Spiking Neurons as an Entropy-Aware Tokenizer for Large Language Models*  
Authors: *Ming Chen, Wenyao Li, Chao Liang, Shi Gu, Peng Lin, De Ma, Huajin Tang, Qian Zheng, Gang Pan*

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?

*Our work proposes a novel tokenization method. It is a fundamental NLP technique focused on improving the efficiency and adaptability of token representation, without introducing new capabilities (e.g., content generation, decision-making, or data collection). The method operates on standard, publicly available benchmark datasets (e.g., XNLI, Wiki40B) and does not involve sensitive data, real-world deployment, or direct human interaction. Therefore, it entails no significant ethical, societal, or security risks beyond those inherent to standard language modeling research.*

### 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 did not explicitly discuss this in the paper. However, all datasets used are standard public benchmarks in multilingual NLP, sourced from publicly available text, which have been widely used in prior work without reported privacy or offensive content concerns.*

#### 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 did not report dataset statistics (e.g., number of examples, train/dev/test splits) in the paper, as we use standard public benchmarks with established splits commonly adopted in the multilingual NLP literature. Full data processing details and split configurations are provided in our anonymized code repository to ensure reproducibility.*

### C. Did you run computational experiments?

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

*Section 4 Experimental Setup; Appendix C; Table 7*

*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.*

- 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 Experimental Setup; All Tables.*

**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.?

*(left blank)*

- 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 did not include a statement about AI assistant use, as they were only employed for code writing and grammatical proofreading tasks that do not involve the generation of scientific claims, methodology, or experimental analysis. The authors retain full responsibility for all intellectual content in the paper.*