

Responsible NLP Checklist

Paper title: *LoPT: Lossless Parallel Tokenization Acceleration for Long Context Inference of Large Language Model*

Authors: *Wei Shao, Zheng Lingchao, Pengyu Wang, Peizhen Zheng, Li Jun, Yuwei Fan*

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, LoPT (Lossless Parallel Tokenization), is fundamentally a correctness-focused solution designed to eliminate the primary risk inherent in prior parallel tokenization methodsnamely, inconsistent or incorrect tokenization due to boundary artifacts. Since our method is formally proven and empirically validated to produce outputs identical to sequential tokenization, there is no novel risk introduced by our approach itself. The risks associated with prior methods (e.g., merging errors) are resolved by our design, so discussing additional risks would not be applicable. Our contributions are centered on performance optimization without compromising correctness, ensuring the framework is risk-neutral relative to the standard tokenization process.

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?

(left blank)

- 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 1 and section 4

C. Did you run computational experiments?

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

section 4 and appendix

- 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?

Time and resources do not allow us to conduct multiple large-scale experiments.

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.?

(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?

The AI assistant was used only for grammar checking and minor rephrasing; such use is considered common writing aid and does not affect scientific content, hence no disclosure.