

Responsible NLP Checklist

Paper title: *Evolving Sparsity: Leveraging Token Importance Dynamics for Efficient LLM Decoding with Sparse Attention*

Authors: *Ruizi Han, Miao Zhang, Ziyue Qiao, Liqiang Nie*

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?

This work focuses on fundamental algorithmic improvements and performance evaluation on standard, publicly available datasets. It does not involve human subjects, sensitive data, or applications that carry foreseeable safety or ethical risks.

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?

The study relies entirely on established, publicly available benchmark datasets that have been previously vetted and anonymized by their creators. As we did not collect new data from individuals, a separate discussion on PII protection steps was not included.

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?

The datasets used are standard research benchmarks with fixed train/test/dev splits widely known in the community. Basic statistics such as the number of examples are consistent with the original versions and are briefly mentioned in our Experimental Setup (Section 5.1) rather than a dedicated section.

C. Did you run computational experiments?

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

section 5.1, section A.1, section A.2

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 used greedy decoding for all experiments to ensure deterministic and reproducible outputs. Since

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.

greedy sampling yields the same result for every run under identical conditions, reporting descriptive statistics such as error bars or mean/variance from multiple runs is not applicable.

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 did not involve human participants or annotators. All data used are from pre-existing, publicly available datasets, thus no instructions were given to participants.

- 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)?

No human participants or crowdworkers were recruited or paid for this study, as all experiments were conducted on established benchmark datasets.

- 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)?

Not applicable. This research utilizes existing public datasets where consent was already obtained by the original data creators according to their respective protocols.

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

This study does not involve new human data collection or human subjects research, and therefore did not require approval from an ethics review board.

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?

In the preparation of this manuscript, we used large language models (LLMs), such as ChatGPT, solely for language polishing and improving the clarity and readability of the text. All scientific content, experimental design, data analysis, and conclusions were independently developed by the authors. The LLM was not involved in any scientific decision-making or data processing.