

Responsible NLP Checklist

Paper title: *OCP: Outlier-Centric Probing for Dynamic Structured Pruning of LLMs*

Authors: *Yang Ji, Ying Sun*

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 improving the efficiency of large language models through structured pruning. It does not introduce new downstream applications or alter the fundamental capabilities of the models in a way that introduces novel ethical or societal risks beyond those already inherent to the foundational models used (e.g., LLaMA, OPT).

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?

This work does not involve the collection or use of new datasets containing personally identifiable information. All experiments are conducted on standard, publicly available benchmarks (e.g., WikiText-2, PTB), where data vetting and anonymization have been handled by the original dataset creators.

- 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?

This work does not introduce new datasets and relies on standard publicly available benchmarks. Detailed dataset statistics are available in their respective original papers, which are properly cited in Section 4.1. Specific details regarding data sampling and sequence lengths for our experiments are provided in Appendix B.

C. Did you run computational experiments?

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

Experimental setup and hyperparameters are described in the main text and/or appendix, including model configurations, pruning ratios, and specific algorithmic parameters (e.g., decay rate, adjustment strength). Please refer to Section 4.1 and Appendix B for full details.

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?

Yes, we report the mean performance averaged over three random seeds for our experiments, as explicitly stated in Section 4.1. Additionally, the standard deviation across these runs is provided alongside the mean values in our detailed result tables (e.g., Table 2 in the main text, and Tables 7-9 in the Appendix).

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 used Claude solely for language polishing.