Responsible NLP Checklist

Paper title: *Dropping Experts, Recombining Neurons: Retraining-Free Pruning for Sparse Mixture-of-Experts LLMs*
Authors: *Yixiao Zhou, Ziyu Zhao, Dongzhou Cheng, zhiliang wu, Jie Gui, Yi Yang, Fei Wu, Yu Cheng, Hehe 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 research introduces DERN, a foundational algorithmic framework for pruning and reconstructing SMoE models to enhance computational efficiency. The nature of this work is purely methodological, focusing on the architectural manipulation of neural networks to reduce resource consumption. It does not engage with user data, specific downstream tasks, or deployment scenarios where direct societal risks (such as bias amplification, malicious use, or privacy violations) typically arise. Any such risks are attributes of the base LLM being compressed, not the compression method itself. On the contrary, by making large models more efficient and accessible, our work can contribute positively to democratizing AI and reducing its environmental footprint. Given this context, a specific section on potential risks was deemed not necessary.

☑ B. Did you use or create scientific artifacts? (e.g. code, datasets, models)

☑ B1. Did you cite the creators of artifacts you used?
See Section 4.1 (Experimental Settings) and Appendix C (Link to Models and Datasets).

☒ B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
Our paper does not explicitly detail the licenses of the artifacts used. However, all models, datasets, and frameworks are from public, reputable sources. We provide direct citations and links to these artifacts in Section 4.1 and Appendix C, which allow readers to find the original license information. This is a common practice for foundational research of this nature.

☒ B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
Our use of existing artifacts, such as public benchmarks and pretrained models, is fully consistent with their intended purpose for academic research. For our created artifact (the implementation of DERN), the intended use is to facilitate the replication and extension of our work within the research

community. While this was not explicitly discussed in the manuscript, our usage and intended purpose fully align with standard academic norms. We provide all necessary citations for readers to verify the terms of the original artifacts.

☒ 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?
Our study utilizes well-established and publicly available academic datasets (e.g., C4, MMLU, and other standard benchmarks). These datasets are widely used by the research community and have undergone prior curation and cleaning processes by their original creators. We relied on these standard, pre-processed versions for our experiments and did not perform additional checks for personally identifying information or offensive content. Therefore, a discussion of such steps is not included in our paper.

☑ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
Yes, documentation for the artifacts used is provided in Section 4.1 (Experimental Settings), as well as in Appendix A and B.

☑ 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?
in Section 4.1

☑ **C. Did you run computational experiments?**

☑ C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
See Section 4.1 (Experimental Settings) and Appendix A (Implementation Details).

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
in Section 4.3 (Ablation Study), Appendix A and B

☑ 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?
in Section 4.2

☑ C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation, such as NLTK, SpaCy, ROUGE, etc.), did you report the implementation, model, and parameter settings used?
in Appendix A and B

☒ **D. Did you use human annotators (e.g., crowdworkers) or research with human subjects?**

N/A 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.?
Our research did not involve human participants or annotators.

N/A 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)?
Our research did not involve recruiting or paying human participants.

N/A 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)?

Our research uses pre-existing public datasets and did not involve collecting new data from human subjects.

☒ N/A D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
We did not collect data; we used existing public datasets. Therefore, an ethics board review for a data collection protocol was not applicable to our study.

☒ D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
Our work uses established public benchmarks. We did not report the demographics of their original annotators as our focus is on the algorithmic method.

☑ **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 an AI assistant only for assistance with proofreading and improving writing clarity. The authors take full responsibility for all final content and its accuracy.