Responsible NLP Checklist

Paper title: *Logits-Based Block Pruning with Affine Transformations for Large Language Models*
Authors: *Zekun Hu, Yichu Xu, De-Chuan Zhan*

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 presents a method for model compression and does not involve the development of a new application or the collection of new datasets. We are not aware of any specific societal risks that are unique to or directly amplified by our technical contributions beyond those common to existing large language models. Therefore, a separate section on potential risks was not included.

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

☑ B1. Did you cite the creators of artifacts you used?
4

☒ B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
The licenses for the artifacts we used are governed by their original providers. For the artifacts we created, we will apply standard open-source licenses upon public release, but this licensing information was not included in the manuscript itself.

☒ 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 is consistent with their intended research purposes as defined by their creators. For the artifacts we created, they are intended for academic research and reproducibility. While our usage aligns with these purposes, our manuscript does not contain an explicit, standalone discussion on this specific topic.

☒ 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 work does not involve collecting or curating any new datasets. We exclusively use existing, widely adopted public benchmarks and pre-trained models for which these concerns have already been addressed by the original creators.

☒ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
Our work is methodological and focuses on model compression techniques. The artifacts we create and use are intended for research reproducibility and are evaluated on established benchmarks whose documentation is provided by their original creators. Therefore, a dedicated documentation section for these domains and demographics within our manuscript was not provided.

☑ 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?
4

☑ 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?
4

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
4

☒ 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?
Due to the computational expense of experiments with large language models, we report results from a single run for each experimental setting. The deterministic nature of our method ensures that results are reproducible.

☑ 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?
4

☒ 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.?
(left blank)

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)?
(left blank)

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)?
(left blank)

N/A D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
(left blank)

N/A D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
(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 an AI assistant during the writing phase of this research to improve the clarity and fluency of the English text.