Responsible NLP Checklist

Paper title: *Rank-Awareness and Angular Constraints: A New Perspective on Learning Sentence Embeddings from NLI Data*
Authors: *Zicheng Zhou, Min Huang, Qinghai Miao*

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?
Yes, potential risks are discussed in the Ethics Statement and Limitations sections on page 9.

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

☑ B1. Did you cite the creators of artifacts you used?
Throughout the paper, particularly in Section 3.1, Section 4.1, Appendix A, Appendix B.

☒ B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
The licenses for the publicly available datasets and models used are standard in the research community.

☒ 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)?
The existing artifacts (e.g., SNLI, MNLI, pre-trained models) were used in line with their standard research purposes. The artifacts created (RAOE models, Enhanced NLI Dataset) are intended for research in sentence representation and related NLP tasks. Our Ethics Statement (page 9) touches upon responsible use and deployment considerations.

☒ 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?
We utilized publicly available and widely used research datasets (SNLI, MNLI, STS benchmarks). While our Ethics Statement (page 9) acknowledges potential inherent societal biases in these datasets, we did not perform additional checks for PII or offensive content beyond the original curation processes of these standard datasets, nor did we apply further anonymization to the source data itself. Our focus was on leveraging these datasets as commonly used in the field.

☑ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
Details of the datasets used (SNLI, MNLI) are provided in Section 3.1 with citations to original documentation. The creation of our "Enhanced NLI Dataset" is documented in Appendix A. The RAOE model architecture, training, and components are described in Section 3 and Appendix B. Performance characteristics across various tasks (STS, SentEval, MTEB) are documented in Section 4.

☑ 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?
Appendices A and B provide statistical details for the datasets used in this study.

☑ **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?
Model parameter counts for various backbones are reported in Table 1. Appendix B details Budget.

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
The experimental setup is detailed in Section 4.1. Hyperparameter settings are provided in Appendix 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?
Descriptive Statistics are reported in Section 4 and Appendix B.

☑ 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?
Appendix B details Parameters For Packages.

☒ **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.?
N/A. Our research did not involve direct interaction with human subjects or new data annotation by human annotators; we utilized existing publicly available datasets.

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)?
N/A. Our research did not involve recruitment or payment of human subjects or annotators.

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)?
N/A. We utilized existing, publicly available datasets (SNLI, MNLI, STS benchmarks). Consent and ethical considerations related to their original collection are addressed by the creators of those datasets.

N/A D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
N/A. Our research utilized existing, publicly available datasets (SNLI, MNLI, STS benchmarks). We did not conduct new data collection from human subjects that would require ethics review board approval for this specific work.

☒ D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?

No. We used existing public datasets (SNLI, MNLI, STS benchmarks). While the original publications for these datasets may provide some annotator details, our paper does not reiterate or analyze these characteristics.

☑ **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?

No. AI-powered tools were utilized for standard grammar checking and code debugging assistance during the preparation of this manuscript, but specific details of their use or the names of the tools were not explicitly included in the submitted paper. These tools were used to improve clarity and correctness but did not contribute to the core research ideas or experimental results.