Responsible NLP Checklist

Paper title: *Benchmarking LLMs on Semantic Overlap Summarization*
Authors: *John Salvador, Naman Bansal, Mousumi Akter, Souvika Sarkar, Anupam Das, Santu Karmaker*

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?
This research is entirely computational and involves only publicly available text data. It does not involve human subjects, personal data, or any procedures that could cause physical, psychological, or societal harm. Therefore, we believe it poses no foreseeable risks to individuals or communities.

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

☑ B1. Did you cite the creators of artifacts you used?
2

N/A B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
(left blank)

☒ 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 artifacts we produced are entirely derived from publicly available datasets that are already released for unrestricted research use. No proprietary or restricted-access data were used, and our artifacts remain within the same open research context.

☒ 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 dataset consists solely of publicly available privacy-policy documents. These are institutional policy texts, not personal communications, and therefore do not contain personally identifying information or offensive content. No anonymization was required.

☒ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
The artifacts are entirely drawn from publicly available privacy-policy datasets that are already thoroughly documented by their original providers; we used them without modification and relied on the original documentation.

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.

☑ 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?
2

☑ **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?
3

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
3

☑ 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?
4

☑ 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?
3

☑ **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.?
A.4

☑ 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)?
3

☒ 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)?
The dataset does not contain any sensitive information

☒ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
the work uses only publicly available privacy-policy documents and does not involve collecting private information from human participants or annotators. Therefore, no ethics review board approval was required.

☐ N/A D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
Demographic and geographic characteristics of the annotator population is irrelevant to the task.

☑ **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?
Writing assistance was for creating outlines and was heavily edited after the fact