Responsible NLP Checklist

Paper title: *Retrieval Augmented Generation based context discovery for ASR*
Authors: *Siskos Dimitrios, Stavros Papadopoulos, Pablo Peso Parada, Jisi Zhang, Karthikeyan Saravanan, Anastasios Drosou*

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 work focuses on evaluating context integration methods for ASR in controlled experimental settings using publicly available datasets (TED-LIUMv3 and Earnings21). It does not involve user data collection, model fine-tuning on sensitive content, or deployment-facing systems. As such, it does not pose immediate ethical, social, or safety risks requiring discussion.

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

☑ B1. Did you cite the creators of artifacts you used?
Sections 2.1, 2.2, 3.1, and 3.2

☒ B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
All artifacts used (MiniLM, TED-LIUMv3, Earnings21, SPGISpeech, LLaMA3.2, and SpeechBrain) are publicly available under research or non-commercial licenses. Since the paper does not involve redistribution or modification of these resources, license terms were not discussed explicitly.

☒ 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)?
All artifacts used in this work (MiniLM, TED-LIUMv3, Earnings21, SPGISpeech, LLaMA3.2, and SpeechBrain) were employed strictly for research purposes, consistent with their intended academic use. As the usage aligned with original licensing and no derivative data was redistributed or deployed beyond research contexts, an explicit discussion was deemed unnecessary.

☒ 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 datasets used in this study (TED-LIUMv3, Earnings21 and SPGISpeech) are publicly available and curated for research, with transcripts that do not include personally identifying information or

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.

offensive content. No additional data collection was performed, and all analysis was conducted on pre-existing, anonymized data

☒ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
he study uses well-documented, publicly available datasets (TED-LIUMv3, Earnings21 and SPGISpeech), for which domain, language, and demographic information is already provided by the dataset creators. As no new artifacts or data were created or modified, additional documentation was not necessary.

☑ 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?
section 3.1

☑ **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?
The ASR model is based on a publicly described architecture (cited in Section 3.2) with a custom implementation following their contextual biasing mechanism. Inference was performed using publicly available models (MiniLM and LLaMA3.2) and a segment-wise ASR pipeline implemented with SpeechBrain. As the work focuses on plug-and-play inference without retraining, detailed compute metrics were not recorded.

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Section 3 and 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?
Section 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?
Common NLP libraries (e.g., NLTK for stopword removal and WordNet definitions, FAISS for nearest-neighbor search) were used with default parameters unless otherwise specified.

☒ **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?
AI assistance was used for paraphrasing or polishing the authors original content but not for technical content, coding, or data analysis. As the contribution of the AI assistant was limited to language refinement and formatting suggestions, it was not included in the main paper body.