Responsible NLP Checklist

Paper title: *Understanding the Modality Gap: An Empirical Study on the Speech-Text Alignment Mechanism of Large Speech Language Models*
Authors: *Bajian Xiang, Shuaijiang Zhao, Tingwei Guo, Wei zou*

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?
The paper focuses on empirical analysis of modality alignment in large speech language models. It does not introduce new models, data collection, or deployment, and does not pose significant safety, ethical, or misuse risks beyond those already established for large language models.

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

☑ B1. Did you cite the creators of artifacts you used?
The creators of all used scientific artifacts (including Whisper-large-v3, LLaMA3.x, Qwen2.5, VoiceBench, and other datasets or models) are cited in Section 2 (Related Work), Section 3.1 (Model Architecture), and in the References.

☒ B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
The models and datasets used in this work (such as Whisper-large-v3, LLaMA, Qwen2, and VoiceBench) are all publicly available under their respective licenses. We only used these artifacts for research purposes in compliance with their terms, and did not redistribute or modify them. Therefore, we did not explicitly discuss their licenses in the paper.

☒ 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 intended use of all artifacts is for research and non-commercial purposes, which matches the context of our work. No artifacts were used or distributed in a way inconsistent with their original terms.

☒ 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?
All the data used in this work (such as VoiceBench and other synthetic datasets) are publicly available

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.

and were either created for research purposes or released under terms prohibiting personally identifying information or offensive content. We did not collect new data from human subjects, and all used datasets are assumed to have been properly anonymized by their original creators. Therefore, no additional steps for anonymization or content filtering were discussed in the paper.

☑ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
Yes. The documentation for the newly created dataset, including data collection procedures, language coverage, dialogue patterns, and filtering mechanisms (for safety, semantic clarity, and naturalness), is provided in Section 3.2 (Experiment Setups

☑ 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?
Relevant statistics for the datasets used and created, including the number of samples, data splits, and total hours, are reported in Section 3.2 (Experiment Setups).

☑ **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?
Section 3.1 (Model Architecture) and Section 3.2 (Experiment Setups).

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Section 3.1 (Model Architecture) and Section 3.2 (Experiment Setups).

☑ 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 3.3 (Results and Analysis)

☑ 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?
Section 3.2 (Experiment Setups)

☒ **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?
The use of AI assistants was limited to minor editing and was not explicitly documented in the paper.