

Responsible NLP Checklist

Paper title: *VCB Bench: An Evaluation Benchmark for Audio-Grounded Large Language Model Conversational Agents*

Authors: *Jiliang Hu, Wenfu Wang, Zuchao Li, Chenxing Li, Yiyang Zhao, Hanzhao Li, Liqiang Zhang, Meng Yu, Dong Yu*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- 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?
(left blank)

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

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?
Although we do not mention it in the paper, we ensure that personally identifiable or offensive content is filtered out when manually processing the data.

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, Appendix A.1

C. Did you run computational experiments?

C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
(left blank)

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

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.?
Appendix A.3.2

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. ACL 2026 used a subset of ARR checklist form.

- 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)?

Internal employees involved in text data processing do not receive additional compensation. For external crowdsourced annotators (from Biaobei (Qingdao) Technology Co., Ltd.) involved in audio recording, compensation was provided based on recording duration, at rates not lower than the local minimum wage.

- 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)?

We do not discuss this in the paper. However, all participants involved in data recording signed written informed consent forms, explicitly agreeing that their voices and transcriptions could be used for academic research and publicly released after anonymization.

- D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?

Our institute does not require formal ethics board approval for non-intrusive, low-risk conversational data collection at the time of collection. Nevertheless, we strictly adhered to our institutional ethical guidelines, ensuring participant informed consent, data anonymization, and that no sensitive topics were involved.

- 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 use AI solely as a writing aid to polish the paper and improve its quality.