

Responsible NLP Checklist

Paper title: *MTR-DuplexBench: Towards a Comprehensive Evaluation of Multi-Round Conversations for Full-Duplex Speech Language Models*

Authors: *Zhang He, Wenqian Cui, Haoning Xu, Xiao-Hui Li, Lei Zhu, Haoli Bai, Ma shaohua, Irwin King*

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?

Appendix E

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?

Appendix E

- 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 3.2

C. Did you run computational experiments?

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

We did not perform any hyperparameter search. All hyperparameters were fixed to standard/default values as reported in the paper.

- 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?

Appendix D. We only performed a single run for each experiment due to computational constraints (or deterministic nature of the method). Therefore, we do not report error bars or summary statistics across multiple runs. The reported results are from that single run.

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.

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 B

- 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)?

All human participants in this study are student volunteers, which do not involve any payment.

- 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)?

For our data, we used two sources. The first is an existing open-source dataset, for which proper licenses and usage permissions were obtained as specified by the dataset provider. The second is synthetic data generated by our own procedure; no human subjects or personal information are involved in the synthetic data. Therefore, consent issues do not apply to the synthetic part, and for the open-source part, we followed the original dataset's terms of use."

- ^{N/A} D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?

No ethics review board approval was sought because our study did not involve any human subjects or collection of personal data. We used only (1) publicly available open-source datasets, which are released under licenses that permit research use, and (2) synthetic data generated by our own procedure, which contains no real human information. Therefore, the data collection protocol is exempt from ethics review.

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 only use AI assistants for the language polish purpose.