Responsible NLP Checklist

Paper title: *Seeing is Believing: Emotion-Aware Audio-Visual Language Modeling for Expressive Speech Generation*

Authors: *Weiting Tan, Jiachen Lian, Hirofumi Inaguma, Paden Tomasello, Philipp Koehn, Xutai Ma*

> 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?
Please refer to our "Limitations and Broader Impact" section

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

☑ B1. Did you cite the creators of artifacts you used?
It's cited in sectoin 5.1 and 5.3 when datasets we used (IEMOCAP, LRS3) are mentioned.

☑ B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
They are mentioned in Ethical Consideration under our "Limitations and Broader Impact" section

☑ 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)?
They are mentioned in Ethical Consideration under our "Limitations and Broader Impact" section

☒ 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?
Our data is from open-sourced video dataset that has already annonymized identities

☒ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
We use publicly accessible datasets and cite them so that their documentation can be found in their original paper.

☑ 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?
it's mentioned in section 5.1 as well as Appendix A

☑ **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?
We report out infrasture and implementation setup (with training hyperparameters) in Appendix D

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
It's mentioned in section 5.1 and Appendix D

☒ 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?
We run every training experiment once due to their high computation cost. We perform different evaluations to validate the effectiveness of our method

☑ 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?
We document our experiment setup in both section 5.1 & 5.3's implemtation section as well as Appendix D

☒ **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.?
We do not use human annotation

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)?
We do not recruit annotators

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)?
we do not employ human work force to collect/curate data

N/A D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
we do not have such data collection process

N/A D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
We do not use annotators for this study

☑ **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?
They are mentioned in Ethical Consideration under our "Limitations and Broader Impact" section