

Dear Reviewers,

Regarding our short-paper submission titled “Crafting a Responsible Dialog System for Collaborative Learning Environments, we have addressed the following reviews:

*It incorporates a dialog policy and controllable response generation(CRG) model to produce reasonable response. Dialog state and actions are determined by consulting with domain expert. The idea is good. It will be interesting to see the result when the work is finished. Clarity should be increased.*

To increase clarity, we revised the manuscript with a more detailed figure caption and further elaboration on our proposed models and architecture.

*Details of utterance classification model and CRG model used to develop the system is not provided. A study method is discussed in the "current study design" section, it mentioned it involves students aged 12 to 17. Is there any particular reason to select this range? Survey result is not discussed in the paper. No of students are also not mentioned.*

We modified the caption for Figure 1 to specify that following ASR transcriptions, the sentences are annotated with utterance-level features that are fed as input to the dialog policy. The utterance classification model currently supported is a BERT [1]-based sequence pair classification Collaborative Problem Solving (CPS) model [2] but we intend to expand this list as more utterance-level classifiers become available for our dataset.

For the controllable response generation model, we have updated a few sentences at the end of section on “Responsible Dialog System Design” to clarify how our response generation model will work. We also specify that in future work we will evaluate different pre-trained models for their viability in the classroom domain when fine-tuned on our lab study data.

The age range of 12-17 is due to the nature of the Jigsaw task [3], which is a part of larger sensor immersion curriculum [4] and was added in collaboration with the NSF National AI Institute for Student-AI Teaming (iSAT). The sensor immersion curriculum is typically implemented at the middle- and high-school levels. We have added these relevant cites to the paper.

Preliminary survey results were excluded due to space constraints and lack of significant results at this stage. We have added our target number of small groups (n =25 groups) we hope to collect during this phase of the study .

Thank you for your consideration,

Margaret Perkoff, Emily Doherty, Jeffrey Bush, and Leanne Hirshfield

- [1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv [cs.CL]*, Oct. 11, 2018. [Online]. Available: <http://arxiv.org/abs/1810.04805>
- [2] Rachel Dickler, Peter Foltz, Nikhil Krishnaswamy, Jacob Whitehill, John Weatherly, Michal Bodzianowski, Maggie Perkoff, Rosy Southwell, Samuel Pugh, Jeff Bush, Michael Chang, Leanne Hirshfield, Daeja Showers, Ananya Ganesh, Zeqian Li, Elita Danilyuk, Xinlu He, Ibrahim Khalil Khebour, Indrani Dey, et al., "iSAT Speech-Based AI Display for Small Group Collaboration in Classrooms," presented at the Interactive event at the Artificial Intelligence in Education Conference (AIED 2022),
- [3] J. Cao, R. Dickler, M. Grace, J. B. Bush, and A. Roncone, "Designing an AI Partner for Jigsaw classrooms," *aichildinteraction.github.io*, 2023, [Online]. Available: [https://aichildinteraction.github.io/preprint/AIIC23\\_paper\\_7399.pdf](https://aichildinteraction.github.io/preprint/AIIC23_paper_7399.pdf)
- [4] Q. Bidy *et al.*, "A professional development model to integrate computational thinking into middle school science through codesigned storylines," *Contemporary issues in technology and teacher education*, vol. 21, no. 1, pp. 53–96, 2021.