International Society of the Learning Sciences

Comparing the Science Talk of AI and Human Students

Alex Barrett, Florida State University, Abarrett3@fsu.edu
Nuodi Zhang, Florida State University, nzhang4@fsu.edu
Fengfeng Ke, Florida State University, fke@admin.fsu.edu
Chih-Pu Dai, University of Hawai'i at Mānoa, cdai@hawaii.edu

Abstract: This poster reports on an exploratory comparison of middle school science classroom discourse from AI-powered virtual student agents and human students. Transcripts from both simulated science classes with preservice teachers and AI students and recordings of real science classes were coded using a framework of student science talk moves. Results suggest that the AI and human discourse is mostly similar, although the AI tended to ask questions much more frequently than human students did.

Introduction and theoretical background

The advent of generative artificial intelligence (AI) has afforded a wealth of research opportunities in the learning sciences. One such opportunity is in simulation-based learning wherein AI-powered virtual agents can facilitate active, situated, and personalized learning (Dai & Ke, 2022). This can be particularly useful for preservice teacher training, a dynamic and performance-oriented profession that has traditionally relied on costly and inauthentic training methods such as peer-to-peer roleplay (Spencer et al., 2019). Despite its potential in overcoming these shortcomings, the efficacy of simulating students using AI for preservice teacher training is underexplored.

Generative AI allows for the development of conversational agents that can be fine-tuned to produce natural language based on supplied text (Dai & Ke, 2022). Therefore, it is theoretically possible to create an AI-powered student agent by fine tuning an AI using a corpus of human student discourse. However, it is unknown whether the discourse produced by AI students in preservice teacher training simulations will match the authenticity of human students'. Therefore, the purpose of this poster is to make exploratory comparisons of AI-powered student discourse with the discourse of human students on which the AI was fine tuned. In doing this we aim to better understand the potential of using generative AI-powered student agents for preservice teacher training.

Method

This study adopted a case-study design and recruited eight preservice STEM teachers to engage with a 3D virtual world-based teacher training simulation with AI-integrated students. Each training session lasted for a minimum of two hours and included at least one hour where participants delivered a lesson to the AI-powered students. The student model was built with OpenAI's generative pretrained transformer 2 (GPT-2) large language model and further fine-tuned with a 24,000-word transcript obtained from open-source recordings of middle school science classrooms (ambitiousscienceteaching.org/video-series/).

Text-based interactions between the preservice teachers and the AI-powered students resulted in 2,431 dialogic turns at talk. These data were coded by two researchers using an adapted framework for analyzing student classroom science talk (Barns et al., 2022). This framework includes the following talk moves in order of scientific rigor: *silence, question, definition/fact, description/observation, under theorized science explanation,* and *fully theorized science explanation*. The researchers standardized their coding on 20% of the data and discussed coding discrepancies until 100% agreement was achieved then coded the remaining data independently. The researchers similarly coded the transcripts of real-world classes which were used to fine-tune the AI. Frequency comparisons and cluster analyses were conducted to explore the similarity between AI and human student discourse.

Results

Figure 1 depicts the talk move comparisons between the AI and human students. Similarities are evident between the AI and human students' silence, definition/fact, and fully theorized science explanation talk moves, which showed differences of less than 5%. The description/observation talk move was the most frequently observed and showed a difference of 6% between AI and humans, with human students using this talk move slightly more. Humans also used the under theorized science explanation talk move more than AI students did with a difference of 15.7%. The largest difference in science talk moves between AI and humans was in asking questions, with AI using this talk move much more frequently than humans at a 21.8% difference.

Figure 1
AI and Human Student Talk Move Comparisons by Percentage of Total Talk.

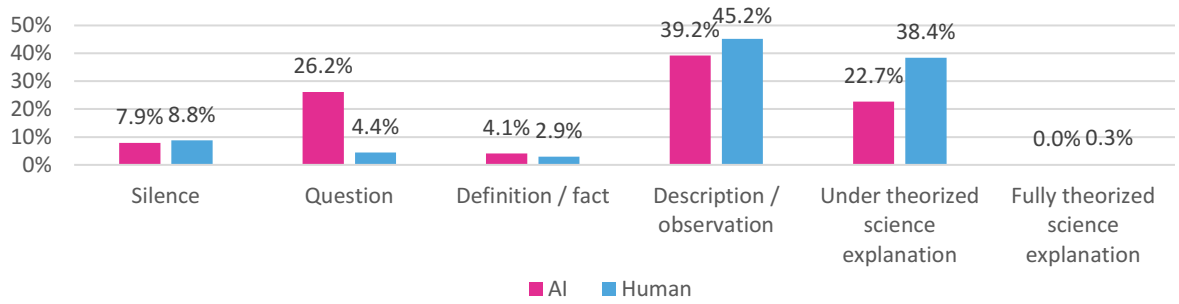
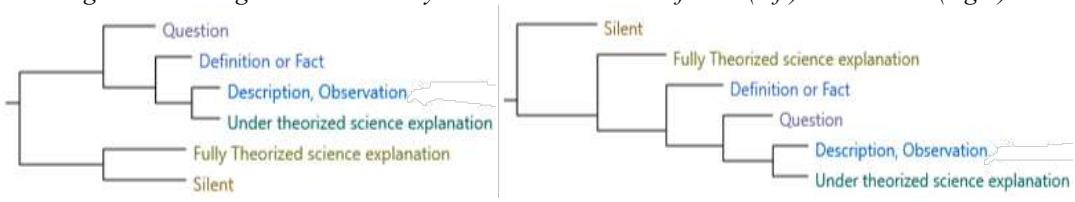


Figure 2 shows the cluster analyses for both the AI and human student talk moves grouped by word similarity. Clustering tended to place talk moves in similar hierarchies for both AI and human students, although the AI dendrogram put silences and fully explained science explanations (which did not occur in the AI data) on a separate branch in a two branch solution, compared to the human dendrogram which shows one cascading branch.

Figure 2
Dendrograms Showing Word Ssimilarity in Coded Talk Moves for AI (left) and Human (right) Students



Discussion and conclusion

Developing AI students for preservice teacher training has the potential for providing critical classroom exposure that is otherwise very difficult to obtain, allowing preservice teachers to transition into real-life classrooms with much more confidence. A critical step toward achieving that goal is to simulate student classroom discourse authentically so that teachers can, in turn, develop discourse abilities that evoke more productive science talk in their students.

This poster reported on an initial exploration of AI-powered student science talk moves in preservice teacher training simulations. Several promising similarities were evident between the AI discourse and its human counterparts, and the cluster analyses saw talk moves grouped correspondingly according to word similarity. However, the AI also digressed from human discourse patterns in key areas, such as question asking and simple science explanations. It is unclear why the AI asked more questions when the data that was used for fine tuning had relatively few. One reason may be that the preservice teacher discourse simply evoked more questions. To better understand the differences between the AI and human student discourse, sequential pattern analyses which include teacher talk moves will be conducted.

References

Barnes, E., Gray, R. & Grinath, A. (2022). Talk moves as pedagogical tools for eliciting and working with student ideas in an undergraduate general biology laboratory. *Science Education, 107*, 89-123. https://doi.org/10.1002/sce.21762

Dai, C.-P. & Ke, F. (2022). Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review. *Computers & Education: Artificial Intelligence*, online. https://doi.org/10.1016/j.caeai.2022.100087

Spencer, S., Drescher, T., Sears, J., Scruggs, A. F. & Schreffler, J. (2019). Comparing the efficacy of virtual simulation to traditional classroom role-play. *Journal of Educational Computing Research, 57*(7), 1772-1785. https://doi.org/10.1177/0735633119855613

Acknowledgments

Research reported in this article is funded by the National Science Foundation, grant 1632965 and grant 2110777.