



## Simulating students: An AI chatbot for teacher training

Saptarshi Bhowmik, Alex Barrett, Fengfeng Ke, Xin Yuan, Sherry Southerland, Chih-Pu Dai, Luke West,  
Zhaihuan Dai

sb17s@my.fsu.edu, abarrett3@fsu.edu, fke@fsu.edu, xyuan@cs.fsu.edu, ssoutherland@admin.fsu.edu,  
cd18m@my.fsu.edu, law19a@my.fsu.edu, zd12@my.fsu.edu  
Florida State University

**Abstract:** This article describes the development of a chatbot designed to simulate students in a 3D virtual environment for the purpose of pre-service teacher training. Using a generative pre-trained transformer-based deep neural network model, researchers created an artificially intelligent chatbot using language resource data from authentic classroom dialogues. Results indicate that the chatbot needs to be fine-tuned with additional programming. This program is intended to be used in future research on teacher-training in virtual simulations.

### Introduction

Virtual 3D environments provide a convenient and immersive space for pre-service teacher training to take place. In these virtual spaces pre-service teachers can practice classroom discourse and classroom management without the pragmatic difficulties of undergoing such training in actual classroom environments with real students (Dieker et al., 2014). Presently, most 3D virtual teacher-training environments rely heavily on puppeteered student avatars to avoid the complexity of natural language processing associated with machine generated discourse (e.g., Cohen et al., 2020). Puppeteered avatars detract from the authenticity of the dialogue and can also be resource intensive because an individual must be employed to play the role of the students. Therefore, this study saw the development of a chatbot which can be used to program student virtual agents to simulate authentic science, technology, engineering, and mathematics (STEM) classroom environments for training purposes.

The contributions of the study are twofold. Firstly, the development of a custom designed chatbot by fine-tuning a pre-trained deep neural network model on classroom conversation data. And secondly, the generation of real-life dialogic responses for domain specific queries using a the chatbot.

### Method

The foundation of the chatbot is a generative pre-trained (GPT-2) transformer-based deep neural network model which was developed by OpenAI (Radford et al., 2019). Pre-training of GPT-2 model was done on a large volume of web data from Reddit making it capable of generative quality text (Fig. 1). For the purpose of the study, the “medium” 355M parameter model (1.5 GB on disk) of GPT-2 was selected, to ensure a good balance between scalability for fine-tuning with large amounts of data and creativity during text generation. The model takes a sequence of tokens as an input and chooses a sequence of output tokens, based on probabilities determined by the inner neural network.

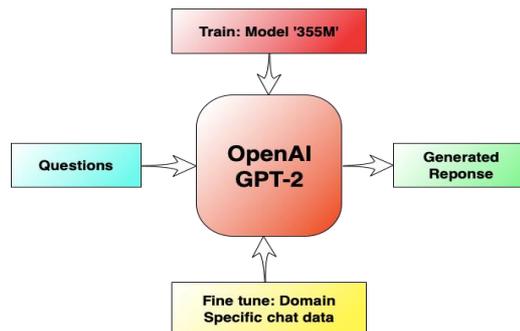
In order for the chatbot to produce dialogue that simulates the diversity and richness of teacher-student interactions in STEM classrooms, data were collected from video recordings of real lessons and manually transcribed. The videos were obtained from online open-access sources such as (<https://rme.org.uk/>) and (<https://ambitioussciencelearning.org/>). Thus far, about 13.5 hours of classroom dialogue have been transcribed, and 9.5 hours have been provided to the chatbot for training. The remaining 4 hours of data were withheld for testing the chatbot.

After the transcription of the dialogue data, it was annotated with tags by two researchers, which guides the chatbot for a more logical and consistent output during response generation. The tags were formatted in a specific order “<Person>\_<Domain>” to maintain singularity as well as preserve scalability for future integration with newer datasets. Once the data format was ready, the chatbot parameters were set for fine-tuning the model. For our study the model was fine-tuned for 500 steps, saving the checkpoints every 100 steps. We also set the temperature parameter to 0.6 to keep a balance between the coherence and randomness of the generated responses. The chatbot is deployed in virtual machine in Google Cloud with 128 GB memory and Nvidia Tesla V100 GPU.

Preliminary testing of the chatbot was done by providing it with authentic queries from the transcribed data withheld for testing purposes. Chatbot responses were assessed subjectively using a linguistics-based metric.

### Figure 1

*Workflow of the designed chatbot*



## Results

The chatbot was inconsistent in its ability to produce appropriate responses. The following example demonstrates the chatbot could return relevant, and appropriately concise answers with the testing data.

*Query: Well that's a very good question. Obviously, how many points do I really need to make a line?*

*Chatbot answer: Two.*

However, the same question met with an inconsistent answer when it was paraphrased. Asking “*What number of points is needed to create a line?*”, the chatbot answered “*Three.*”

Another noted difficulty in the preliminary analysis is that a lack of contextual information forced chatbot responses to be interpreted in isolation. For instance, in the following example the data were taken from a classroom activity in which the teacher was demonstrating the properties of pressure by blowing up a balloon.

*Query: Ok, so what happened? What are your observations?*

*Actual student response: You applied pressure and it got bigger.*

*Chatbot answer: Um, that it's hard.*

Although pragmatically and grammatically, the chatbot answer was not necessarily infelicitous, and indeed might be observed from a student in an actual classroom, the lack of contextual information regarding the balloon makes the chatbot a difficult partner in an ongoing dialogue.

## Discussion and conclusion

Programming a chatbot to accurately portray the verbal behavior of student avatars is an iterative process central to the creation of a classroom simulation which does not rely on puppeteered avatars. One likely reason for the chatbot's answers being inconsistent despite queries aimed at producing identical answers is that the temperature parameter, here set to 0.6, may be too low for the chatbot to interpret queries that stray too far from the original text in the data. Setting this parameter very low will produce only responses found in the dataset and setting it very high may produce incoherent responses. Furthermore, programming the chatbot to incorporate ongoing dialogue into its resources might allow it to show more consistency in its responses by following dialogic threads.

The authors intend to employ the chatbot with virtual agents using *Open Simulator*, which will be used in studies addressing the efficacy of using virtual simulations for pre-service teacher training.

## Acknowledgment

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

## References

- Cohen, J., Wong, V., Krishnamachari, A. & Berlin, R. (2020). Teacher Coaching in a Simulated Environment. *Educational Evaluation and Policy Analysis*, 42(2), 208-231. <https://doi.org/10.3102/0162373720906217>
- Dieker, L., Rodriguez, J., Lignugaris-Kraft, B., Hynes, M. & Hughes, C. (2014). The Potential of Simulated Environments in Teacher Education: Current and Future Possibilities. *Teacher Education and Special Education*, 37(1), 21-33. <https://doi.org/10.1177/0888406413512683>
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 1-24.