Shih, JL. et al. (Eds.) (2023). Proceedings of the 31st International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Toward a Virtual Human Exhibit for Public AI Education

Ning WANG[a*], Timothy HURT[b], Ari KRAKOWSKI[b], Eric GREENWALD[b] , Omkar MASUR[a], Boxi FU[a] & Chirag MERCHENT[a]
[a]*Institute for Creative Technologies, University of Southern California, USA*
[b]*Lawrence Hall of Science, University of California, Berkeley, USA*
*nwang@ict.usc.edu

Abstract: It is important to prepare the world's youth and the future workforce with fundamental knowledge of Artificial Intelligence (AI). Informal settings, such as museums, offer great opportunities in helping youth and the general public learn about AI. In this paper, we will discuss the design of a Virtual Human Exhibit that aims to communicate to the public about the capabilities and impact of AI through AI technologies used in virtual humans.

Keywords: K-12 AI education, learning in informal settings, virtual human

1. Introduction

Artificial Intelligence (AI) is a foundational technology that is transforming our work and daily lives. It is critical to prepare future generations with basic knowledge of AI. It is also critical for parents to learn about AI, as they are often the decision-makers for their children's use of AI. Informal settings, such as museums, offer unique opportunities to reach out to two-generation learners – parent and child in learning about AI. In this paper, we discuss the design of a Virtual Human Exhibit that aims to communicate to the public about the capabilities and impact of AI through AI technologies used in virtual humans. The exhibit is design for science museums with visitors between 5- to 12-year-old, accompanied by their parents or guardians. The exhibit currently includes a number of activities that center around how virtual humans utilizes AI to realize functionalities similar to humans, such as recognize facial expressions. While the exhibit is still expanding through research and development, part of the exhibit has already been piloted at working museums with hundreds of visitors.

2. Virtual Human Exhibit

The Virtual Human Exhibit includes a number of interactive activities where the visitors can walk up, experience the AI-driven technologies used to help a virtual human character think, feel, speak, and act like a real human, and reflect upon the implications on their own lives. The current implementation of the exhibit includes *1-2-3 Smile!* and *Mirror Mirror*, two activities where visitors interact with a couple of facial expression recognition AIs; *Virtual Comedian*, where the virtual human tells joke and use multimodal behavior analysis (e.g., from vision and speech) to infer if the visitors like the joke; *Sensing Showcase*, where virtual human converse with the visitors while the user interface displays visualizations of data processing behind-the-scenes (e.g., head pose and facial landmark tracking; speech recognition); and *Silly Face*, an activity discussed in detail in the following section.

Unlike learning taking place in formal settings, such as classrooms, a key challenge in designing educational experiences in informal settings, such as museum, is *engagement*. In this free-choice environments with many exhibits to choose from, visitors often spend little time deciding if they want to interact with an exhibit. This means that the design of the exhibit should (1) attract the visitors' attention and pique their interest to interact with it. Given that the target learner population is 5- to 12-year-old children and their accompanying guardians, we designed the virtual human to be a character of similar age to the young visitors. Unlike

formal learning settings, with teachers to scaffold the learning and answer question, (2) the learning experience in a museum needs to be self-explainable, without the need for human assistance or intervention. Given that our visitors are primarily 5- to 12-year-old children, with limited reading ability, we designed the instructions used in the exhibit in simple languages. The instructions are given verbally instead of as text displayed on the screen. While the AIs driving the integrated virtual human are complex, explaining how such complex AI systems work is no small feat either, particularly to an audience of young age, without engineering background, in a short amount of time in a museum setting. We have developed a D-I-R (detect-interpret-respond) framework to illustrate the high-level process of how AI behind virtual humans work to serve as the underlying structure for scaffolding the understanding of such process (Greenwald, Krakowski, Hurt, & Wang, 2023).

One of the first exhibit activities that has gone through iterations of design, implementation, and user testing is *Silly Face*. In this activity, the visitors help the virtual human learn to recognize their silly and serious faces. The activity is organized in five phases:

- In the initial **Introduction** phase, the virtual human greets the visitors as they walk up to the exhibit (e.g., "*Hi there! Would you like to try an A.I. activity with me? Press the red button to start!*" and briefly introduces the task (e.g., "*For this activity, I am going to use A.I. to learn more about your facial expressions. Here, let me zoom in to your face.*"). This phase clearly marks the beginning of the activity. In our earlier testing when this phase was not included, the young visitors were unclear about which phase the exhibit was in and what they should do to start it.
- In the **Data Collection** phase, the virtual human guides the visitors to pose silly and serious faces and captures images of those faces as training dataset (Figure 1, Top-Left). The virtual human provides the young visitors with clear instructions, examples, and encouragement to help them throughout this phase. The virtual human first provides an example silly face image: "*Do you know how to make a silly face? Here's what my silly face looks like...*", then asks the visitors to post silly faces: "*I would like to learn what your silly face looks like. Could you make some silly faces to help me learn?*" During the data capturing, the virtual human continues to speak to the visitors: "*Keep being silly; I learn better if I can see all of your different silly expressions*." Such phrases serve not only as words of encouragements, but also as indications that the capturing phase is still ongoing. At the end of silly face images capturing, the virtual human concludes this round of capturing and introduces the next round: "*Wow, I really like your silly faces! I think I have enough data to learn what your silly face looks like. I'd like to learn another one of your expressions. Here is my serious face. Can you make serious faces?*"
- In the **Model Construction** phase, the virtual human learns from the images captured and builds a classifier to recognize the visitors' silly and serious expressions. In addition to communicating verbally to the visitors, such as "*Let me take a moment to learn the difference between your serious faces and silly faces.*", the user interface also plays animations where individual pictures are highlighted to one by one, to simulate that the virtual human is examining and learning from each image.
- In the **Model Testing** phase, the virtual human tries to recognize the visitors' facial expressions using the learned model: "*Try making a bunch of different silly and serious faces to see if I am good at determining which one you are making.*" (Figure 1 Top-center and Top-Right). This phase lasts around 30 seconds.
- In the final **Reflection** phase, the virtual human reviews the images recognized as silly and serious faces with high confidence levels (Figure 1 Bottom-left and Bottom-center): "*I am pretty confident that you were being silly in this photo? Do you agree? Does this photo look similar to the silly faces you showed me earlier?*" and the ones with low confidence level (e.g., Figure 1 Bottom-Right): "*I had a hard time figuring out if you were making a serious face or a silly face in this picture? Why do you think I struggled?*". The goal of the reviews is to help the visitors reflect how training data impacts the AI performances, and the limitation of the state-of-the-art AI (e.g., hard to identify facial expressions not seeing in the training dataset).

This activity captures the visitor inputs via a single physical button, which allows the visitors to start (or restart) the activity, and a webcam, which captures the visitors' facial expressions. MediaPipe is used for face recognition tracking (Lugaresi et al., 2019). The facial expression classifier is implemented using MobileNetV3 (Koonce & Koonce, 2021). The virtual human's speech is synthesized in real-time using Microsoft text-to-speech. The speech gesture and lip-synch are procedurally generated in real-time through NVBG (Lee and Marsella, 2006) and Cerebella (Lhommet et al., 2015) based on text of the speech.

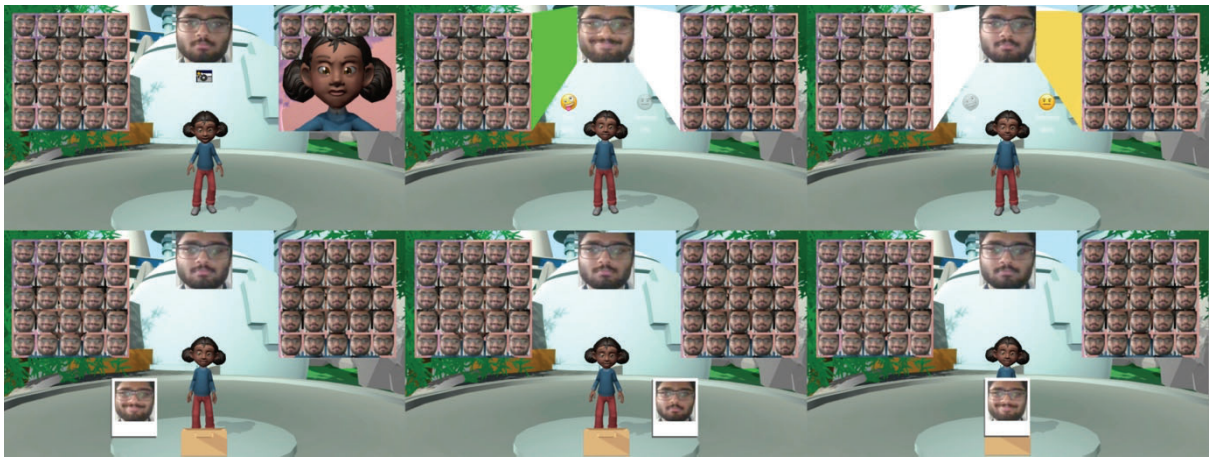

Figure 1. Screen captures of the different phases in the *Silly Face* activity of the Virtual Human Exhibit. *Top Left*: the virtual human gathers images of a visitor's silly and serious faces as training data. *Top Middle and Right*: after using the training data to build a silly/serious expression recognizer, the visitor tests how well its performances. *Bottom Row*: the virtual human reviews the recognized expressions with the visitor. Note: a member of the research team posed as the visitor in these illustrations.

3. Discussion

The *Silly Face* activity has recently studied through a 9-day exhibit at a science museum in the United States, with over 500 visitors. Observations from the project's third-party independent evaluation team indicate that the exhibit created high levels of visitor engagement and dwell time, compared to other exhibits at the museum. Other activities from the exhibit, such as *Mirror Mirror* and *Virtual Comedian* have been piloted during science fairs at local high schools. Studies on these activities with target audiences at science museums are underway to gather data on age-appropriateness of these activities (e.g., can the young museum visitors understand the jokes told by the virtual human comedian?) and the guided interactions to help visitors grasps the AI behind virtual humans.

Acknowledgements

This research was supported by the United States National Science Foundation under Grant # 2116109. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Greenwald, E., Krakowski, A., Hurt, T., & Wang, N. (2023). Detect-Interpret-Respond: A framework to ground the design of student inquiry into AI systems. *Proceedings of the AIED Workshop on AI Education in K-12*.

Koonce, B., & Koonce, B. (2021). MobileNetV3. *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*, 125-144.

Lee, J. and Marsella, S. (2006). Nonverbal behavior generator for embodied conversational agents. In Proceedings of Intelligent Virtual Agents, 243–255. Springer.

Lhommet, M., Xu, Y., and Marsella, S. (2015). Cerebella: automatic generation of nonverbal behavior for virtual humans. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.L., Yong M.G., & Lee, J. (2019). Mediapipe: A framework for building perception pipelines.