Ghosts in a Shell: An Immersive Art Experience for ALIFE23

Alyssa M Adams, Oneris Rico, Nicholas Guttenberg, Olaf Witkowski
Cross Labs
Cross Compass
Kyoto, Japan
alyssa.adams@cross-compass.com

Abstract

Visit this link to see a video version of this abstract.

At this moment in technological history, it seems that AI-powered technology has the potential to evolve into almost anything within the next 20 years. While we expect machines to don various forms of intelligence, we also expect to integrate them into our daily lives in ways we haven't yet imagined. How will their presence and capabilities affect our everyday human experience? While we're often (rightfully) thinking about how our day-to-day lives will change, we rarely pause to consider the experience of the *machines* themselves. But there's a good reason for this. What a machine "experiences" is difficult to define, much less measure. We also have difficulty understanding the concept of experience *in general*. We don't fully understand the experiences of the many other living creatures who've shared our world for millennia. So while we cannot yet measure how models like ChatGPT[1] or Stable Diffusion[2] experience a written conversation, we may be able to experiment with different ways of translating a machine "experience" to a human one. How do current algorithms translate their inputs into an output, and what happens along the way? In this art installation, we introduce wearable technology meant to translate aspects of what a trained model allocates attention to into something a human can experience.

Transformer Models and Attention

Transformer models are some of the most popular algorithms making headlines today. Many popular models that generate images, videos, text, and audio are all specifically tuned transformers. Their ability to generate is due to how transformer architectures are modified by training on data. This data comprises millions or billions of human-generated examples of inputs to outputs. As a result, the trained model can then take an input and produce an output similar to something it has "seen" in the data.

The transformer architecture is powerful and unique because it uses extra layers to implement a form of attention. These layers specifically encode relationships between particular elements in the data set. This allows the model to pick up on patterns of elements that occur together more frequently than those that don't.

In our installation, we pick out the attention layers in trained models and map them onto something humans can understand while the models process input. In other words, our installation allows humans to experience the attention layers of trained transformer models in real-time to translate aspects of machine "experience" into a human-machine hybrid experience. The technology interface allows human agents to experience their immediate environment in a new way–one that is enhanced by transformer attention models.

Translating machine attention into a human experience

We don't wish to eliminate human autonomy from this immersive art experience; we feel that AI technology should empower the agency and autonomy of humans. AI technology does not suggest actions for human users to take but provides additional/different information about the immediate environment. With this new information, the human agent can decide how to interact with the environment. At ALIFE23, someone from the art installation team will present an introduction to the installation and demonstrate how people can experience it. In addition, a text description will be available to illustrate the connection between this technological experience and what we experience when we hold a seashell up to our ear. In both cases, the shape of an object (the shell and the wearable technology) changes our sensory experience when we interact with the environment.

People who wear the immersive helmet receive visual and audio input that reflects the active attention layers from the AI models. This gives the wearer a sense of space that is *enhanced* by AI models in addition to their own biological hardware (such as eyes and ears). The wearer has full bodily autonomy and can interact with their surroundings as they usually would while wearing the helmet–the only difference is that they are *experiencing* the world from inputs based on the attention layers of trained transformer models.

Below is a figure that shows the design of the immersive wearable technology:

A visual preview of what a person will experience while wearing this technology can be found at this link. This video

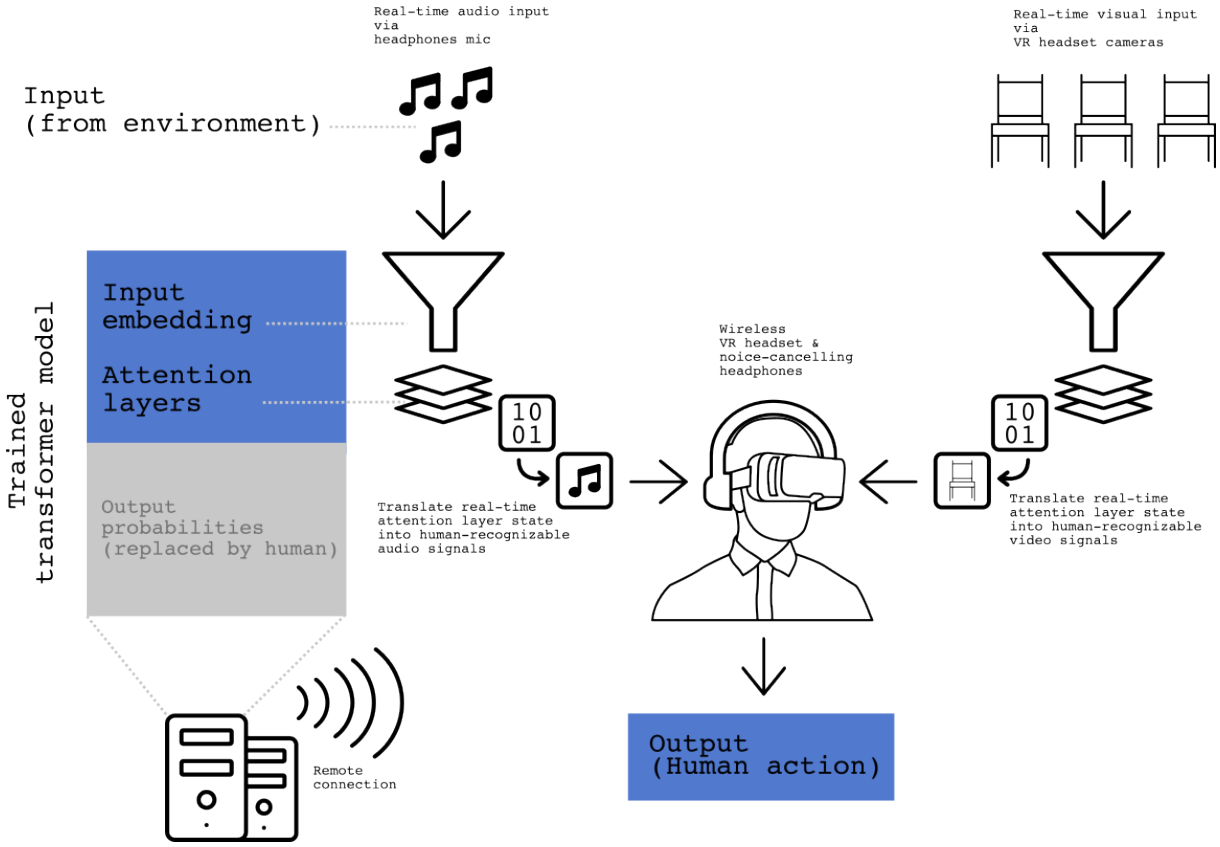

Figure 1: A diagram showing how parts of transformer models (from inputs to attention layers) are translated into something a human can hear and see. A headset and noise-canceling headphones will provide the immersion experience. Icons were provided by Noun Project and are attributed to the following artists: rdesign, Amiryshakiel, sapon prayetno, Icon Lauk, Ribbla Team, Agni, Travis Avery, Dan Stack, Icon Depot

demonstrates the live video feed participants will see as they wear the technology. The visual field is obstructed except for areas the attention layers are actively processing. This way, the participants can only see the areas of the visual field that the attention layers "see." A similar mechanism will be used to process incoming auditory signals in the environment.

What we can learn

Humans are already equipped with tools to experience the world around them, but this method allows us to fuse additional layers onto this experience. In particular, these modifications are designed to translate machine attention– the "experience" of popular generative algorithms (abstract weights and vectors)–into a medium understandable by human biological machinery (eyes and ears).

We are particularly interested in how this experience makes people feel. What does it feel like to directly see the different kinds of "ghosts" (visual and audio artifacts) that machines create and "experience"? Does it cause humans to be more empathetic and give moral consideration to machines? After each participant finishes this experience, we will ask them how they feel with an audio recorder so they can describe their experience naturally. We hope these insights will lend us further insights into the manufactured experiences of multiple agents, whether human, machine, or a mix of both.

References

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2021.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc., 2020.