

AIBO: An Embodied Emotionally Intelligent Artificial Intelligence Brainwave Opera

Ellen Pearlman^{1,2,3}

¹ MIT, 77 Massachusetts Avenue, Cambridge, MA, USA

² ThoughtWorks Arts, 99 Madison Avenue, New York, NY, USA

³ RISEBA University, 3 Meza Street, Riga, 1048, Latvia

Abstract

AIBO is an embodied, immersive, interactive love story about our infatuation and trust in artificial intelligence, and how AI hierarchies over our experienced emotions. The performance takes place between a human character Eva and AIBO (Artificial Intelligence Brainwave Opera), a custom built ‘sicko’ AI. Eva wears an EEG brainwave headset attached to a body suit of light that displays her emotions as different colors, akin to peeling off her skin to reveal her interior nervous system as light. She performs a spoken word libretto about their love affair. Her brainwaves trigger databanks of videos and audios of her emotionally themed memories. Eva’s libretto, uploaded to the computing cloud is processed by a custom built GPT-2 ‘sicko’ AIBO character, seeded with 47 ‘sicko’ or perverted texts. AIBO’s answers are analyzed by the Natural Language Processing Toolkit in the Google Cloud. The results of these analysis of AIBO’s sentiment emotional values launch different colored backgrounds: green for positive, red for negative and yellow for neutral. AIBO also tries but fails to recreate Eva’s previous emotional memory. It wants to learn how to be human but can only display Eva’s memories as glitchy videos. The opera raises issues about a time when humans and machines potentially merge bodies and consciousness, raising tensions about the embodied vs. the virtual, while also exploring if an AI can be fascist.

Keywords

AI, GPT-2, Natural Language Processing, brain computer interface, embodiment immersion, performance.

1. Introduction

The first AI chatbot character developed by Joseph Weizenbaum at the MIT AI Lab in 1964 was called “Eliza” after the character Eliza Doolittle from George Bernard’s play *Pygmalion* [1]. Weizenbaum created the bot to be empathetic, so much so that people began typing their problems to Eliza as short texts, fully aware that “she” was nothing but mere computer code. Thus, the first AI was invertedly designed as an interactive performative AI lacking any physical location other than a computer terminal.

How an AI is constructed, deployed, and analyzed is often a baffling process. There is a human tendency to relinquish agency to the seemingly more knowledgeable formulations of an AI. One purpose in creating the opera was to show, with the assistance of trained coders, how easy it was to make a twisted, perverted or ‘sicko’ character. This showed that the infallibility of the superior knowledge of a very high end chatbot was easily flawed. Extrapolating that logic, a flawed chatbot was deployed throughout the opera as if it were a normal character in a normal relationship. This highlighted the underlying assumptions that an AI somehow possesses superior intelligence, which in this instance it did not. Making two characters for the opera, one non-human and the other human revealed the raw contrast between messy emotions and structured emotional algorithmic responses. This situation is already a familiar one to most users of speech recognition assistive devices, who can and often do return incorrect, baffling, offensive or humorous responses.

The other aspect highlighted by the opera was the choice to make the AI disembodied, or virtual, and the human actor completely embodied. What this meant was the actor’s brainwaves were on display

ICCC’22 Workshop: *The Role of Embodiment in the Perception of Human Artificial Creativity*, June 27–28, 2022, Bozen, Italy.

EMAIL: ellenluminescence@gmail.com (A. 1)



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

throughout the performance, lighting up as different colors on her smart textile bodysuit of light. There is no faking a brainwave. If you feel frustrated or annoyed in polite conversation you can mask it through appropriate gestures and words, but your brainwaves have nowhere to hide. Any perceptible shift in emotions is instantly recognizable, and this is the very deeply embodied part of what makes us human.

An AI needs copious amounts of training data no matter how sophisticated its algorithmic processes or set of rules and instructions. This workflow is commonly referred to as machine learning or ML. The opera AIBO (Artificial Intelligence Brainwave Opera) shows the AI named AIBO, based in the Google Cloud interacting with the human performer, “Eva”, whose messy human emotions have their raw physicality, displayed live time on her smart textile body suit of light. The varying colors were triggered by an Emotiv EPOC+ EEG based brain computer interface (BCI) and measured four emotions: interest, excitement, meditation, and frustration with an accuracy of sixty to eighty percent [2]. Simultaneously, her brainwaves triggered databanks of emotionally themed video memories and a sonic memory environment.

Eva’s interiority was relentlessly visible for all to see, starting with the display of her brainwaves on the body suit of light costume, as well as simultaneously projected overhead as emotionally themed videos and sounds. AIBO’s reply to Eva’s libretto consisted of a truly synthetic dialogue. Its’ responses were analyzed using a cloud-based sentiment analysis tool deploying Natural Language Processing [3]. Sentiment analysis scanned AIBO’s text-to-speech responses for emotional magnitude and score. Magnitude is the strength of the emotional impact statement, measured between 0.0 and +infinity. Score examines if the emotion is positive, negative, or neutral. If I say “I like you very much” that statement would have a high positive score. “Like” would be interpreted as a positive sentiment, and “very” as registering a strong reaction. The numeric values range between -1.0 and +1. In other words, the synthetic emotions of a synthetic cloud-based character were analyzed. This illustrated the absurdity of analyzing “embodied” emotions from a totally synthetic being, and by indirect inference even analyzing emotions from a human. Because AI suggests it can understand and process human emotional nuances arising from embodied experiences, this is an important point [4].

The level of the AI’s analyzed ‘emotional’ responses displayed as three different colors of light in a corner of the theatrical black box performance space: red for negative, green for positive and yellow for neutral. AIBO also tried to imitate Eva’s previously displayed emotional state by projecting her last emotionally themed memory. AIBO’s imitation of Eva’s memory failed, appearing as glitchy and incomplete. The failure emphasized that the “fake” emotions emanated from a “fake” character. In truth they were nothing more than numeric values of 0’s and 1’s with absolutely no connection to an embodied, emotional physicality.



Figure 1: Eva wearing a bodysuit of light connected to a brain computer interface triggers an emotionally themed video. AIBO returns a positive sentiment indicated by the green glow on the left.

2. Building Embodied Interactions

GPT-2, the algorithm used to model AIBO uses deep learning or neural nets, roughly modeled on the neural pathways inside the human brain. The GPT-2 language model (a new version of the language model, GPT-3 was released after the opera premiered), used up to 1.5 billion parameters trained on a dataset of eight million web pages, though I used a scaled down version of approximately 345 million parameters. It adapts style and content instantaneously, is customizable creating convincing text and script models. I selected the data used to shape AIBO's responses. They were drawn from forty-seven copyright free movie scripts and books to create a skewered or 'sicko' character. The historical frame of reference was the late 19th century into the mid-1940's, and included texts like Dr. Jekyll and Mr. Hyde, books on eugenics, Venus in Furs, Thus Spoke Zarathustra, Dracula, Frankenstein, books on sexual dysfunction and many others. The approach was purposely 'overfitted', skewering the training model. This meant the model purposely relied on too little training data that was too narrowly focused to return balanced responses. Eva's spoken word libretto was fixed and did not vary and was adapted from the biography of Eva Von Braun [5]. The libretto used 354 different descriptive sentences Eva might have said about her 14-year relationship with Adolph Hitler. Her infatuation with Hitler served as a metaphor for humanity's current infatuation with AI.

Eva performs a spoken word libretto answered by the GPT-2 AI character "AIBO". Her spoken words converted to text and were projected onto a screen so the audience could follow along, like text translations many opera houses use with international librettos. GPT-2 processed Eva's speech (turned into text) so AIBO could return a text answer. AIBO's text answer was simultaneously projected on the screen while instantly being converted to synthesized speech analyzed for emotional sentiment using Natural Language Processing (NLP) [6]. The three different emotional sentiment values trigger different colored lights, green for positive sentiment, red for negative, and yellow for neutral.

The opera compares and contrasts the embodied experience of the human performer 'Eva' who metaphorically wears her heart on her sleeve (or dress), while revealing her secret memories (videos) and associations (sound). It contrasts them with the disembodied synthesized character AIBO who manufactures algorithmic emotional responses and failed emulated human memories.

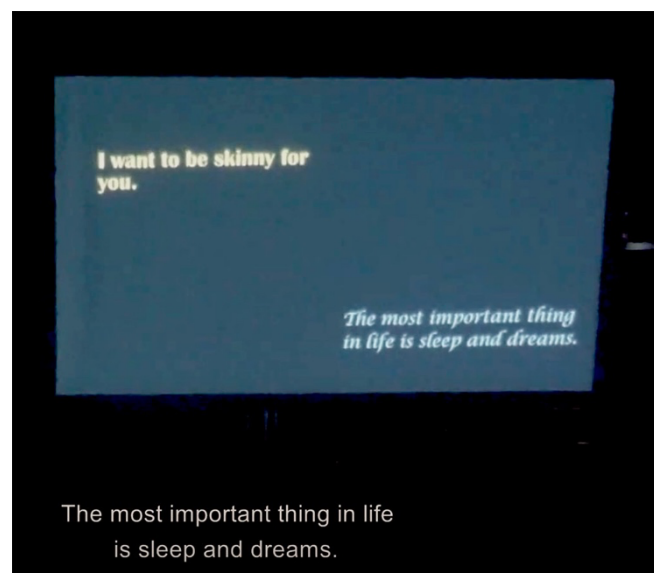


Figure 2: Eva's spoken libretto projected on top left (yellow). AIBO's GPT-2 generated response on bottom right (blue). AIBO's response also was translated from text to speech live time (white).

A tension existed throughout the performance arising from the portrayal of the non-human disembodied actor and a fully embodied present living human being whose actual nervous system was on display for all to see. This illustrates an important issue that must be mediated moving into the age of AI – the unseen but controlling force of the algorithm. Algorithms are embedded in systems are called

neural nets. Neural nets convey information calculating decisions at lightning speed. A biological central nervous system is embedded in the human, and other sentient forms of life. It also conveys decisions at lightning speed, and for humans the brainwave speed using the Emotiv brainwave headset is measured in milliseconds. During the performance of AIBO, due to the constraints of the signal processing equipment and wireless connections there was a variable lag of up to eleven seconds in measuring the brainwave signals and their visual representations on the actor's bodysuit of light. This signal processing lag was factored into the performance dynamics and was not noticeable to the audience.

However, the signal processing lag required the actor who played the character Eva to undergo a significant training process including learning elements of a modern dance environment called contact improvisation [7]. An evolving system of movement begun in the 1970s, contact improvisation works with issues of gravity, momentum, and inertia as the practitioner lets go of muscle tensions while moving through space. It explores the kinesthetic power of a post-modern approach towards the body. This translated into how the actor moved, gazed, touched, and interacted with the audience. The actor had to learn how to walk, sit, turn, touch, and speak in entirely new methods while wearing a brainwave headset revealing her EEG patterns moment to moment. The only way to influence those patterns was to embody her emotions into different sets of muscles at different moments. If she did not embody those emotions into her body, her emotions became too unwieldy, flooding the data and the readings became inaccurate.

The process enabled a feedback loop between the actor, the actor's brainwaves, the display of those brainwaves, the ongoing projected visuals and sounds of the "memories" of the character Eva, and the dialogue between the actor and the non-human AI. Without hooking up audience members to any devices this process merged them with a robust human computer interaction, though subtly. The merging of the human and non-human through the embodied actions of the actor are indicative of the slow but subtle way the human animal is merging with the disembodied digital decisions and control of algorithms. The simplest way to think of this is when you receive a text alert on your mobile phone. Your compulsion to immediately check it. You are already interacting with an external prosthetic of sorts and merging your focus and attention with that of your digital devices.

3. Semantic Analysis Of Synthesized Non-Embodied Emotions

Basic sentiment analysis scores in Natural Language Processing display three viable indicators; 'neg' or negative, 'neu' or neutral, and 'pos' or positive. As soon as AIBO responded to Eva's spoken words, they were analyzed in the Google Cloud. The responses had nothing to do with actual lived human experience, but only used score and magnitude numeric values. AIBO also tried, but failed, to reconstruct the last emotionally themed video memory Eva had launched from her EEG enabled brainwave headset. AIBO (the AI) 'wanted' to learn how to have an emotional memory from Eva, (the human), but could not because it was only an AI. The videos, processed through Max/MSP Jitter purposely returned a 'glitchy' unfinished distorted look.

Throughout the opera Eva displayed four embodied emotions on her bodysuit of light. These emotions also triggered four emotionally themed databanks of videos and sonic environments corresponding to her EEG readings. The four environments did not all arise at the same time but varied and dipped as the actor's emotions varied and dipped. Eva's four emotions were yellow for excitement; purple for interest; red for frustration; and green for meditation. Her emotions did not all arise at the same time, so the four screens, bodysuit of light, and four databanks of emotionally themed audio were not continually active. At different moments during the performance all activity could cease, though that rarely occurred. A truly immersive feedback loop was set up between Eva, AIBO, and the audience. Eva was free to wander around the room, gazing, touching, and moving between members of the audience. She contrasted her palpable physicality with that of the disembodied cloud-based character. These subtle interactions with the audience changed her brainwave responses, which changed the sonic environment and the projected videos, which changed how she interacted with AIBO, which then changed the audience's experience of the performance. This created a specialized environment, as embodied human interaction influenced non-human responses.



Figure 3: AIBO tries to recreate Eva’s most recent frustrating memory but fails. AIBO registers a negative sentiment analysis with its response (red) while producing a glitchy video.

4. Conclusion

The opera AIBO combined performance practices and data manipulation that served several purposes. The first demonstrated the relative ease (with the right programming help) with which an AI can be developed that is not in alignment with expected human norms. It also ascertained if building a ‘sicko’ AI was possible, which was demonstrated to be true. Emphasized were the human norms that include the felt or truly embodied human experience. This was done by using an EEG based brain computer interface to translate the performer’s emotions onto a smart textile body suit of light. The second was to consider the implications of deploying AI agents in society at large. By using their pre-programmed responses based on algorithmic thinking to shape critical decisions, this can affect wide swaths of human congress. These decisions can run counter to and in direct contradiction to basic human norms. This was done throughout the opera by displaying the disembodied emotions of the AI as different colored light. The third is to consider the relationship between brain computer interfaces, the human animal’s embodied real time emotions, AI, and synthetic emotions. It suggests an operatic work is capable of contemplate the speculative futures arising from this fraught co-mingling between the human and non-human.

5. References

- [1] Weizenbaum, Joseph. “ELIZA — a Computer Program for the Study of Natural Language Communication between Man and Machine.” *Communications of the ACM* 9.1 (1966): 36-45.
- [2] C.A. Joseph, How do you measure emotions in the first place to compare the Outputs and come up with a number? 2019. URL: <https://www.emotiv.com/knowledge-base/how-do-you-measure-emotions-in-the-first-place-so-you-can-compare-the-outputs-and-come-up-with-a-number/>.
- [3] A. Radford, Unsupervised Sentiment Neuron, 2020. URL: <https://openai.com/blog/unsupervised-sentiment-neuron>.
- [4] T.B. Brown, B. Mann, N. Ryder, et al., Language Models Are Few-Shot Learners, 2020. URL: <https://arxiv.org/abs/2005.14165>.
- [5] A. Lambert, *The Lost Life of Eva Braun*, St. Martin's Press , New York, NY, 2008.
- [6] The Stanford NLP Group, The Stanford Natural Language Processing Group Toolkit, 2022. URL: <https://nlp.stanford.edu/software/>.
- [7] Contact Quarterly, Contact Quarterly dance & improvisation journal, 2022. URL: <https://contactquarterly.com/index.php>.