

Ellen Pearlman

***How to Build a
'Sicko' AI. AIBO:
An Emotionally
Intelligent
Artificial
Intelligence
AI Brainwave***

ADAMarts

Volume 4, 2023

Architecture and Design

DOI: 10.32025/adamarts23007

Received: 14.08.2023

Accepted: 15.09.2023

Abstract

Using the GPT algorithm originally developed by OpenAI in February 2019, a skewed or ‘sicko’ AI character was created as a real-time entity running in the Google cloud. The algorithm allows imitations of human dialogue that produce fake and often realistic interactions emanating from computer cloud-based agents. The character was created as one of two characters in the emotionally intelligent artificial intelligence brainwave opera AIBO (Artificial Intelligence Brainwave Opera). The spoken word opera rhetorically inquired “Can an AI be fascist?” and “Can an AI have epigenetic or inherited traumatic memory?” through the interplay of human and non-human characters. This contribution discusses certain aspects involved in creating the GPT cloud-based character AIBO.

Keywords

GPT, natural language processing, interactive performance, brainwave opera

...

Introduction

AIBO (Artificial Intelligence Brainwave Opera) is an interactive spoken word new media opera that depicts two characters: Eva, a human performer wearing a brain–computer interface attached to a body suit of light that displays four emotions in real time – frustration, interest, excitement, and meditation – and AIBO, a custom-built GPT-2 AI residing in the Google Cloud, as GPT-2 was the only version available at the time in the public domain. AIBO is based on the biography of the young, naïve Eva Braun’s 14-year-long relationship with her perverted, sadistic lover Adolph Hitler, represented by the AI (Lambert, 2008). Artificial intelligence has come a long way since Joseph Weizenbaum developed the first AI chatbot at the MIT AI Lab in 1964, naming the program “Eliza” after the character Eliza Doolittle from George Bernard Shaw’s play *Pygmalion* (Weizenbaum, 1966). He created it to be empathetic, so much so that people began typing out their problems to Eliza, fully cognizant that “she” was mere computer code. AIBO, a spoken word opera, used a Google cloud-based AI for the character AIBO, who interacted with the human character Eva in real time throughout the performance. Besides using Open AI’s GPT-2 algorithm, which is a predictive algorithm, AIBO also deployed Google’s cloud-based sentiment analysis function in the AI character’s answers as part of the dramatic interplay. A predictive algorithm predicts the next word in a sequence of words, and sentiment analysis defines a string of text as either positive, negative, or neutral. The human performer Eva also wore an EEG brain–computer interface attached to a body suit of light that displayed her emotions as they occurred as different colours of light, a visual focal point throughout the actual performance, especially for the audience. The AIBO AI was constructed specifically to be ‘sicko’ or have characteristics that skewed or mimicked a character with twisted emotional and mental problems. The implications



Figure 1. Performer wearing a bodysuit of light showing her four emotions connected to a wireless EEG brain-computer interface. Photo by Ellen Pearlman.

of AIBO and Eva's 'romance' are a metaphor for humans' infatuation with artificial intelligence, and the opera implies an imagined speculative future. The text responses generated by the AI were secondarily processed using the Stanford Natural Language Processing Toolkit (NLPT) and were examined for their magnitude and score (Stanford NLP Group 2014). The 'emotional' analysis, as opposed to the algorithmically generated GPT-2 text-based answers, were displayed as three different colours of light. The light was red for negative, green for positive and yellow for neutral.

The opera therefore deployed synthetic emotions of a synthetic cloud-based character to illustrate just how fraught an algorithmic analysis of emotions can be. "Fake" emotions emanated from a "fake" character. The emotions, in fact, were really nothing more than weighted numeric values. Since AI will be attempting to understand and process human emotional nuance in the future, the opera speculates on issues AI will have dealing with human non-quantifiable experiences. This includes but is not limited to its ability to be fascist, and its ability, or lack thereof, to understand the subtleties of the human condition, such as epigenetic or inherited traumatic memories.

Recipe for the AI

Throughout the evening, the performer who portrayed Eva used a preselected spoken word libretto of 342 sentences relating to her love affair with AIBO, spanning the fourteen years of their love affair. The libretto was based on a biography of Eva Braun, *The Lost Life of Eva Braun*, by Angela Lambert. AIBO focused on Braun's actual fourteen-

Figure 2. AIBO lighting up with the colour green for 'positive' sentiment as Eva's brainwaves trigger a video and audio. Photo by Taavet Jansen.



year relationship with Adolph Hitler. The GPT-2-built ‘sicko’ cloud character was an imagined designed proxy for Hitler.

Eva’s spoken word libretto was converted to text and projected onto a screen so the audience could follow along, akin to the translation scripts accompanying many foreign operas that are projected during a performance. The GPT-2 AIBO processed Eva’s text, sent it to the Google cloud and returned a text answer from AIBO while simultaneously projecting the answer onto the overhead screen. AIBO’s response was instantly converted to synthesized speech so the audience could also hear it, along with the music that was being triggered by various threshold levels of Eva’s brainwaves. The response from AIBO was also simultaneously analyzed for emotional sentiment in the Google cloud using the Natural Language Processing toolkit. The three different emotional sentiment values from AIBO triggered different coloured lights suffusing an area of the performance space: green for positive, red for negative and yellow for neutral sentiment.

Contextualized in terms of the entire performance of AIBO, the GPT-2 database was just one aspect of the performance. It would be too complex to include the brain-computer interface, or the triggering of videos and a sonic environment by Eva, except as they relate to her interaction with the GPT-2 AI character AIBO. Open AI created GPT-2. It predicts the next word of a text if given all the words that came before it (Radford et al., 2015) (Brown et al., 2020). GPT uses deep learning or neural nets, roughly modelled on the neural pathways inside the human brain, more complex than simple algorithmic processing. GPT-3, not available when the opera AIBO was developed, has been updated to be even more convincing than GPT-2 [6]. GPT-2 outperforms state-of-the-art language modelling scores known as “zero-shot” settings from any other known language models currently in use (Radford et al., 2015). AIBO’s data was seeded with copyright-free curated texts to create a skewed or ‘sicko’ AI. This

meant training a complex predictor on too few examples. Included were forty-seven copyright-free movie scripts and books spanning the late 19th century to the mid-1940s, the historical time frame depicted for the character AIBO. These included Dr. Jekyll and Mr. Hyde, Venus in Furs, Thus Spoke Zarathustra, Dracula, Frankenstein, and texts on eugenics, masculinity, and sexual dysfunction. The approach was purposely ‘overfitted’, meaning it had too high a variance and tried too hard to fit the purposely curated training set. This skewed the training model so that it could not generalize competently any types of new examples, an effect that was deliberately built in. Experimenting with the temperature parameter within GPT-2’s construction adjusted the sampling behaviour of the provided texts. Temperature is a float value that controls the randomness of answers. Lower temperature results in less random results and higher temperature results in more random sentence completion. Below is a sample dialogue between EVA (scripted) and AIBO (GPT-2 generated from the cloud). The temperature variations in this example were focused on the parameters 1.2.0 and 1.2.5 respectively.

EVA: I was given my first camera at age 13.

AIBO: So, you threw away your other one.

EVA: The photo store I worked in was only 20 minutes from my home.

AIBO: How well you handled the Third Reich.

EVA: I was a junior assistant in the studio and darkroom.

AIBO: Eight o’clock.

EVA: I typed, filled in invoices, and modelled in the photo studio.

AIBO: How does an old maid get that phony swag?

EVA: My boss took photos of powerful men.

AIBO: Why wasn’t his name Oberführer Kaunitz?

EVA: I was given my first camera at age 13.

AIBO: That’s because you knew that every day, and every evening, your expression wouldn’t become more interesting – or any other day by shooting you.

EVA: The photo store I worked in was only 20 minutes from my home.

AIBO: That said, I've seen several such photographs of you.

EVA: I was a junior assistant in the studio and darkroom.

AIBO: Did I mention you were a woman, as well?

EVA: I typed, filled in invoices, and modelled in the photo studio.

AIBO: Very well, but it is noted that you never typed after that.

EVA: My boss took photos of powerful men.

AIBO: Phenomenal!

Sequence and routing

The information and signal routing for the GPT-2 occurred in real time with almost no lag. It interacted with a live human performer, routed to the Google cloud, transformed from speech to text, and displayed a coloured light against a wall indicating the sentiment analysis of its response. It also triggered a glitchy video that showed AIBO trying but failing to capture the memory of Eva's last emotion. Because AIBO was only an AI, it had no way to truly understand human emotion, which is why the videos were glitchy. The cycle of how the routing worked began when Eva said a line from the predetermined libretto. Any one of the 342 lines could appear as text on a screen, and the performer would repeat them with convincing depth and feeling.

As she spoke, the text was also routed to the GPT-2 (AIBO) in the Google Cloud. AIBO replied instantly, and its reply was projected as text on a screen next to Eva's statement, as well as instantly translated to synthesized speech for the audience to hear. AIBO's response was further analyzed in the Google Cloud for its emotional content, referred to as sentiment analysis – positive, negative, or neutral, using the Natural Language Processing Toolkit. This sentiment analysis triggered a coloured light: positive if green, negative if red, and yellow if neutral. AIBO also

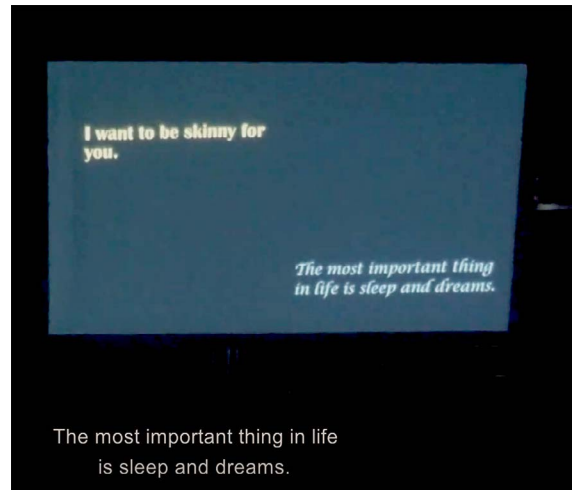


Figure 3. Upper left in yellow – Eva's spoken word libretto; lower right – AIBO's real-time response from the Google Cloud; bottom left – AIBO's text translated to speech. Photo by Ellen Pearlman.

tried to emulate a video of Eva's last emotional memory. That is because AIBO wanted to learn how to become human, and how to feel, sensations that are part of the core of human identity. It took the last emotionally themed video that Eva launched from her brainwave headset and tried its best to match it up with its limited repertoire consisting of only positive, negative, and neutral sentiment. However, AIBO's ability to emulate human feeling was imperfect. The video appeared on a separate monitor placed on the floor as glitchy and distorted because AIBO, being an algorithm, could not really emulate discrete emotions. Eva then spoke the next line in the libretto, and the cycle began anew.

Analyzing sentiment

Analyzing AIBO's emotional sentiment illustrated that completely synthetic algorithmic answers can also return synthetic, but conceptually



Figure 4. AIBO experiencing a negative sentiment (red) and displaying a negative glitchy video (Eva's last negative memory) on a video monitor. Photo by Ellen Pearlman.

“valid” responses. Basic sentiment analysis scores displayed three indicators: ‘neg’ or negative, ‘neu’ or neutral, and ‘pos’ or positive. AIBO tried but failed to reconstruct the most recent emotionally themed video memory that Eva had launched from her EEG-enabled brainwave headset. In the opera, Eva’s bodysuit of light was a specially constructed smart textile costume, akin to an exteriorized nervous system, wirelessly connected to her brain–computer interface. If Eva experienced the emotion of frustration, displayed on her bodysuit of light as red, a video showing a ‘frustrating’ experience from her ‘memory’ was projected. If AIBO’s emotional sentiment from its last answer was negative, it retrieved Eva’s previous frustrating video memory, lit up the background with red light, and displayed the glitchy video on a monitor. Though this seems complex in the context of a performance, it was briefly explained to the audience before the performance commenced. The displays of light, sounds and visuals were understood by the viewer/participants, as evinced by the comments and questions that arose after the performance had ended, during the Q&A session.

Conclusion

AIBO incorporated the GPT-2 algorithm into an artificial character built to be ‘sicko’ or perverted using copyright-free texts centred on human dysfunction, power, fascism, love, perversion, and dominance. The AI character was part of a two-person spoken word opera, with the other character being a human performer. The loosely adapted story illustrated the 14-year relationship between Eva Braun and Adolph Hitler. A basic emotional sentiment analysis was performed on AIBO’s responses using Natural Language Processing. The sentiment analysis was displayed as green (positive), yellow (neutral) and red (negative). The AI also tried to imitate human emotional memories and was intentionally unsuccessful at doing so. This raises the question of whether AI can have human memories and suggests new fields of research in the coming years as AI attempts to become more ‘human’. During each of the two forty-five-minute performances, most of the audience members remained behind after the performance was over to discuss their experience and to ask questions about the societal implications of using artificial intelligence. It was an unexpected but welcome occurrence and illustrates how impactful the questions raised by AIBO were to a general audience. It also demonstrates the relative ease with which an AI can be developed that is not in alignment with expected human norms. The implications of this scenario suggest deploying AI agents in society at large, and using their preprogrammed responses based on algorithmic thinking needs to be reframed in order to shape critical decisions regarding wide swaths of human congress. What is at risk are human feelings being reduced to overquantified and set responses that will automatically categorize and sort large and important parts of future human congress in narrow and potentially harmful ways.

Acknowledgments

This project was funded by a Vertigo STARTS Laureate Grant in collaboration with GoPro Social. Further support was provided by ThoughtWorks Arts.

...



Figure 5: Eva launching negative (red), interested (purple) and meditative (green) (outside of photo) images from her brainwave headset while reciting the libretto to AIBO. Photo by Taavet Jansen.

References

Lambert, A., 2008. *The Lost Life of Eva Braun: A Biography*. St. Martin's Griffin. New York, USA.

Weizenbaum, J., 1966. ELIZA – a computer program for the study of natural language communication between man and machine. *Communications of the ACM* 9, Issue 1, pp. 36-45. Available at: <https://doi.org/10.1145/365153.365168>

Stanford Natural Language Processing Group Software., 2014. *The Stanford Core Natural Language Processing Toolkit*. Available at: <https://nlp.stanford.edu/software/> [Accessed March 1, 2018]

Radford, A. Jozefowicz, R. and Sutskever, I., 2017. *Learning to Generate Reviews and Discovering Sentiment*. Available at: <https://arxiv.org/pdf/1704.01444.pdf> [Accessed July 1, 2019]

Brown, T. B. et al., 2020. *Language Models are Few-Shot Learners*. ArXiv. Available at: <https://arxiv.org/abs/2005.14165> [Accessed June 6, 2020]

OpenAI., 2020. *OpenAI Licenses GPT-3 Technology to Microsoft*. Available at: <https://openai.com/blog/openai-licenses-gpt-3-technology-to-microsoft/> [Accessed March 23, 2021]

