

## **Report on Integration and Enhancement**

### **Model Integration:**

For Milestone 2, I enhanced my project by integrating the sentiment analysis model from ML5.js alongside the custom Teachable Machine model that I developed in Milestone 1. My original model was designed to classify images into two categories, "Pass" or "Kill," based on facial expressions (smiling for "Pass" and frowning for "Kill"). I decided to integrate the sentiment analysis model to add a deeper layer of interactivity and to further demonstrate the arbitrariness of AI decision-making processes.

The sentiment model takes a user-generated text input and evaluates the emotional sentiment behind the text. The final output of my system now combines the confidence score from the image classification and the sentiment score. Depending on the final calculated value, the system makes a decision on whether the user "Passes" or is "Killed." This integration adds complexity and an additional interaction layer by involving user text input along with their facial expression.

### **Enhanced Functionality:**

The key new interaction introduced in this phase is the sentiment analysis that complements the visual classification. The user can type any sentence they wish, and the sentiment analysis will gauge the positivity or negativity of that sentence. This sentiment score is then combined with the confidence score from the Teachable Machine image classifier. The final classification is determined based on these combined values.

For example:

If the label from the Teachable Machine model is "Pass," I add the sentiment score to the model's confidence score. If the result is greater than 1, the user "Passes."

If the label is "Kill," I subtract the confidence from 1 and add the sentiment score. If the result is less than 1, the user is "Killed."

This combination of image and text inputs to determine the final outcome further emphasizes the opaque, arbitrary nature of AI decision-making. It showcases how these systems can be influenced by unrelated data (e.g., a positive sentiment in a text) in ways that may not be rational or justifiable.

### **Challenges and Solutions:**

One of the challenges I faced was finding a meaningful way to combine the two different types of model outputs (image classification confidence and sentiment analysis score) to create a cohesive final decision. I addressed this by experimenting with different mathematical combinations and decided on a system where the sentiment score adjusts the confidence score from the image classifier. This approach maintains the original idea of randomness and arbitrariness in decision-making but adds a new

dimension to user interaction.

Another challenge was managing the user experience, ensuring that the additional complexity introduced by the sentiment model did not overwhelm or confuse the user. I simplified the user interface by adding a button that triggers the sentiment analysis, allowing users to see the result after they have provided their text input.

### **Inspirational Works:**

I continued drawing inspiration from the game *Papers, Please* and the novel *1984*, which served as foundational influences for my project. *Papers, Please* inspired the idea of making arbitrary, bureaucratic decisions with significant consequences, which mirrors the AI's classification process in my project. *1984* contributed the overarching theme of a dystopian society controlled by a biased, powerful algorithm, which I brought to life in the form of the "Big Brother" algorithm.

An additional layer of inspiration emerged when a friend of mine accidentally opened the project on my computer and was shocked by the result. This spontaneous reaction reminded me of the film *M3GAN*, where an AI-driven robot makes unpredictable and frightening decisions. The movie reinforced my goal to explore the eerie unpredictability of AI behavior and how it can evoke shock or fear in users. This experience drove home the idea that AI systems, much like the robot in *M3GAN*, can be perceived as terrifying due to their randomness and unpredictability—further emphasizing the dangers of unaccountable AI decision-making.