Report on Model Training Process

Model Training

For this project, I used **Teachable Machine** to train a model that classifies images into two distinct classes: "Pass" and "Kill." The model’s goal is to simulate an extreme scenario where a future algorithm, referred to as “The Big Brother,” controls life-and-death decisions based on an individual's appearance.

The training data consisted of:

- **"Pass" Class**: A series of images of me smiling.
- **"Kill" Class**: A set of images of me frowning.

The objective of this training was to highlight how the AI's decisions can be entirely arbitrary, and yet due to its opaque, black-box nature, those decisions can still have serious consequences. After training, the model was integrated into a **P5.js** sketch, allowing real-time interaction where users could submit their pictures to see whether they would be classified as “Pass” or “Kill.”

Rationale

I chose this model and features for several reasons:

- **Bias in AI**: The use of a simplistic classification between "Pass" and "Kill" based solely on facial expressions demonstrates how easily an AI can be biased. In this case, smiling leads to survival, while frowning leads to elimination—a completely arbitrary decision-making process.
- **Black-box Nature of AI**: The training process was simple, but the model can be presented as something more complex and inscrutable, which mirrors the reality of many AI systems in use today. The goal was to show how such systems can appear mysterious or authoritative even when they are based on flawed logic or incomplete data.

Challenges and Solutions

One challenge I faced during the model training process was related to handling the background in the images. Initially, I considered adding an additional class, such as "Empty" or "Not Detected," to filter out the background, thinking it might improve the model’s focus on facial expressions. However, I quickly realized that this step was unnecessary. The model’s decisions are arbitrary by design, so it doesn't matter whether the background is present or not.

This realization reinforced the central concept of the project: the model's classifications are ultimately based on subjective and biased criteria. I could have just as easily trained the model using pictures of dogs for the "Pass" class and pictures of cats for the "Kill" class. This highlights the arbitrary nature of AI training and how easily its outputs can be manipulated, regardless of the input data’s relevance.

Inspiration

This project draws inspiration from two significant sources: the game **Papers, Please** and the novel **1984** by George Orwell.

- In **Papers, Please**, players take on the role of a border inspector, responsible for deciding who can enter a dystopian nation. The game emphasizes the weight of bureaucratic decisions and how they can impact human lives, often based on arbitrary criteria. This concept of subjective decision-making heavily influenced the "Pass" or "Kill" mechanic in my project, where an AI algorithm makes similarly arbitrary choices about life and death.
- Orwell's **1984** serves as the thematic backbone of the project, with its depiction of a totalitarian regime governed by "Big Brother." In my project, "The Big Brother" algorithm echoes the book's surveillance state and its ability to control human lives through opaque, biased systems. The AI in my project, like Orwell's Big Brother, is presented as an unquestionable authority, despite being based on flawed, simplistic decision-making processes.