**Inspiration Long Form**

Technology giants such as Microsoft, Apple, and Amazon, plus specialized companies like Kairos and Affectiva are betting big on AI and computers to recognize human emotion. Some estimates indicate these technologies may become an ~$25 billion business by 2023. But how and why will these technologies be used, and as always, is it ethical?

The answer to how and why is relatively simple; humans are emotional beings and we make a lot of decisions based on emotions. And a lot of these decisions involve money. When people are sad; they may indulge in retail therapy. When in in love; might be more inclined to buy rings. The list is endless. But more importantly, if a computer can recognize your individual emotions, whether by your voice talking to your Alexa/google home/etc., or your face staring into your camera and link that emotion to your searches, this creates another effective avenue for companies to market their products to you.

For example, lets imagine a computer can recognize your facial features or voice as being sad, and in the past when displaying similar features/tones you searched for a lot of lounge wear. Before you even realize it, the computer may be able to recognize this pattern and start suggesting you a new pair of sweatpants. Just like cookies, search histories, etc. companies want to use your data and information to personalize advertisements and products to you; and your emotions are just another avenue to accomplish this multi-billion-dollar goal.

In a less capitalist vein however, recognizing human emotions via sound and feature is also an especially important part of AI and robot development. To ever create a reasonable robot analogue of a human, it must be able to at least express a semblance or mimic emotion. Recognizing human patterns and establishing proper responses to those patters is a key part of AI development.

With these technologies, two major questions arise: is it ethical and is it accurate? Human emotions are endlessly complex; even many of us struggle to correctly identify other’s feelings. Smiles can be sarcastic or genuine; a frown can be a joke or even correlate with fear. In the same manner, a tone of sarcasm and other nuances of voice may be unintelligible to a computer. Not to mention the infinite varieties of hand or body gestures that can accompany a facial expression to convey an emotion. Add the fact that emotions can vary across cultures, or even regions; the issue of accurately picking an emotion is a tall task. The best way to navigate this issue so far has simply been to use a massive amount of data points from various mediums.

Of course, the ethics involved are also hot button issue. How do we control what companies or apps can utilize these features? Should computers even be able/allowed to use our emotions or collect data on them? How do we ensure consumer privacy? As with many of these developments, legislation or control over their use may not materialize until after it has already been deployed.

Considering the potential future impact of these technologies, we decided to do create a site that uses machine learning libraries to train and recognize a person's emotion from a voice sample.

**Inspiration Short Form (for actual about page)**

As humans and computers become more and more intertwined both in the workplace and at home, researchers and companies alike are attempting to harness the power of machine learning to recognize human emotion. This has applications in marketing; emotion plays a large role in consumer behaviors/decisions, and is projected to be a $25 billion dollar business by 2023, and in AI/robotics to make robots more responsive and life-like, or even in our own cars to help prevent dangerous driving caused by emotional distress. This is just a small subset of possible applications that computer recognition of emotion can have in our world.

Considering immense potential impact for these recognition technologies, we wanted to create a site that uses a model trained with machine learning libraries to recognize a person’s emotions from a voice audio sample.

**Challenges**

Because the model is trained using 24 actors, the model becomes slightly ‘biased’ or used to those similar voice patterns. Every person in the real world has unique inflections, cadence, and speech patterns so using random/new voices with the trained model may result in some of the discrepancies in prediction. Another factor may be use of words, the model was trained on similar sentence structures and words across voice actors. It is possible that different words, with different sound pronunciation and stress can be confused with a different emotion. I.e. if not a lot of sharp consonants like ‘cart’ or ‘greet’ etc. are used in training, its possible that the increased intensity to produce those sounds may result in an inaccurate emotion output from the model. Another potential source of error is authenticity of emotion. Emotions are extremely complex and result from a multitude of semi-subjective feelings. In many cases, these complex states are hard to conjure up or imitate without accompanying stimuli, resulting in inconsistent results in model training. Think about how difficult it is to be ‘fake mad’ at someone without cracking a smile. Some papers have attempted to use [EEG signals from the CNS](https://www.sciencedirect.com/science/article/pii/S1566253519302532) to predict a person’s emotion more accurately. As discussed in the ‘inspiration’ section, the most reliable way to reduce these possibilities is to use ton’s and ton’s of data; hundreds of actors, varying sentences, regional dialects, etc. to try and create a more holistic model. However, this model alone takes ~2 hours to train each time, so with class time considerations and complexity of the data set, we are proud of the accomplishments.