**Frida, Suhail and Kasturi**

**Class Presentation**

**900 words**

**Prof. Susan McGregor**

**The story**

Aggression Detectors: The Unproven, Invasive Surveillance Technology Schools Are Using to Monitor Students *(Jeff Kao and Jack Gillum, ProPublica, July 2019)*

Story: https://features.propublica.org/aggression-detector/the-unproven-invasive-surveillance-technology-schools-are-using-to-monitor-students/

Link:

Methodology: https://projects.propublica.org/graphics/aggression-detector-data-analysis#fn3

**Its genesis**

Jack Gillum was at a police chief conference in Orlando, Florida when he came across the device and its marketing. He bought one and then investigated its accuracy with Jeff Kao.

The device is made by the Dutch company Sound Intelligence; originally it was meant to be used in hospital wards so patients could signal distress.

**The algorithm design**

The device was manufactured by the Dutch company Sound Intelligence. The reporters were not given access to the demographic composition of the training set to see whether or not it was representative of the school populations it was being used for. Kao said that the When the reporters asked if the data included children recorded at a school, they didn’t receive a response; The device is also marketed to hospitals and prisons, and the reporters believe that the same training data is used for all of these cases.

Data labelling was done manually for audio frames, which are tiny snippets of sound basically devoid of context. When ProPublica did its testing and labelled their sound clips, they had a team of three vote on “aggressive” or “not aggressive” designations to split ties.

The algorithm rates sounds on a 0 to 1 scale of aggression, and will mark sounds above a certain threshold to be aggressive. (The device could also indicate vocal strain or distress, which the reporters also interpreted as a verdict of “aggressive”. given the device’s marketing.) During installation, you can adjust a device’s sensitivity, and hence its threshold; the way that the algorithm weights different features however remains unchanged. Lowering the threshold could decrease the number of false positives but might increase the number false negatives that the algorithm reports.

**Testing**

The reporters made the decision early that they were not going to do a comprehensive and exhaustive study of the device and its applications in schools. The logistical challenges of getting permission to install these devices across schools were too great; rather than try to calculate a false positive rate, the reporters decided to present a few situations in which the device performed poorly, both for logistical reasons and for storytelling purposes. They also decided to play pictionary to have reactions be more genuine.

They modified the device so that they could directly input sound, and then wrote a script to match sound inputs with the aggression labels since there was a slight time lag. They then recorded snippets of audio on a device functionally identical to the microphone in the aggression detection set, manually classified them as aggressive or not and fed them to the device to see if its classifications matched their judgment.

In the LINUX operating system that came with the device, Gillum found a file of features that the algorithm was looking at, and reporters were then able to work backwards to a degree. They ran multiple linear regressions to see which (combinations of) audio features might be contributing to an “aggressive” designation.

**Is the issue the algorithm or its use in schools?**

The reporters concluded that the algorithm was flawed, which could lead to two undesirable outcomes — either the detector would escalate situations because administrators would not have a good understanding of the device’s false positive rate, or the detector would go off so often unnecessarily that it would become “a boy who cried wolf” situation. Off the record, they had an interview with a school administrator who said that device went off all the time. In the story, a school official recounts how singing “happy birthday” set the device off.

As Gillum pointed out, even if the algorithm were perfect, it would still bring up questions around student privacy, surveillance and school discipline that need to be addressed. Since the device doesn’t actually record audio but studies its component parts, the company argues that it doesn’t actually violate privacy. However, the reporters found that the audio files were stored on the system, which “administrators to record, replay and store those snippets of conversation indefinitely.” Societally, are we comfortable with this?

On the subject of school discipline, the reporters couldn’t identify cases where students had been suspended as a result of the algorithm. Nevertheless, such a case is not out of the bounds of possibility, and much like in our discussions about the use of algorithms in the justice system, transparency in decision making is a key component of school discipline. Is a machine learning algorithm really welcome here?

The story also doesn’t investigate whether the detector performs better than human judgment; in that sense, Gillum wondered whether this story was one of a “hammer without a nail.”

**Building on this work**

Kao talked a little bit about wanting to include an interactive component to the story where you could scream into your computer and see what the algorithm would likely do. But given that they could only reverse engineer a replica algorithm that was not identical to that used by Sound Intelligence, they decided it would be misleading.

It’s unclear who exactly has oversight of technology adoption by schools; as far as the reporters were aware, the main statutes that govern this behavior are privacy laws. It would be interesting to see what debates exist on this front especially when you compare video surveillance and the aggression detector algorithms.

Better understanding the bias of the training data would also allow more exploration of bias in implementation.