**What are the ethical problems with making this model?**

I see a few ethical dilemmas with the development of the ‘Gaydar’ algorithm. The first of these is that Dr. Kosinski and Mr. Wang use a utilitarian argument to justify the creation of this algorithm. They say that they created the algorithm to show that it could be done and subsequently draw attention to the privacy risks of such an algorithm. Essentially, they are saying that the end, that attention will be drawn to the privacy risks, will justify the means, the exposure of members of the LGBTQ+ community. These means, however, expose LGBTQ+ people to breaches of privacy in a number of ways, and therefore are unethical. The first breach in privacy comes from the use of web scraping of profile photos of individuals who did not consent to their image being used for this study. The second breach in privacy is that the algorithm could be used to ‘dox’ or ‘out’ people that may want to remain closeted about their sexual identity. Following this argument to its logical conclusion, the algorithm could be used to unfairly target LGBTQ+ individuals. I also found it to be ethically questionable when Dr. Kosinski began to speculate on the implications of the prenatal hormone theory on his study and algorithm. If you use faulty or unagreed upon science to build an argument, your argument is inherently faulty because it was not based in solid science.

**What arguments can you make that the model might be wrong?**

The algorithm could have a number of flaws that would cause straight or gay people to be identified incorrectly. The first of these is that the algorithms were only trained on a specific subset of the population – only American, white men, who were openly gay or presumed to be straight. This selection bias would be evident if this algorithm were used to predict whether a Black or Asian man were gay. The algorithm could make a improper prediction because those groups were not used to train the algorithm. Additionally, because of the nature of neural networks and algorithms of this nature, the authors of the algorithm and paper are unable to specifically identify which features of the face were the features that were used to predict a person’s sexuality. For instance, the article cites another algorithm that found that lighting and photo quality had a strong influence on prediction variables of neural networks.

**Do you think the computer can get the prediction right more often than a human?**

I think the most salient piece of information regarding this question is the following quote from the NYT article: “Dr. Kosinski said that they went to great lengths to guarantee that such confounders did not influence their results. Still, he agreed that it’s easier to teach a machine to see than to understand what it has seen” (Murphy, 2017). Yes, I believe that if given a proper and unbiased set of data, a computer may be able to make more precise and accurate predictions. However, because humans are the ones that write the algorithms, there is inherent bias built into each algorithm – just as Dr. O’Neil outlines in her TED talk, *The Era of Blind Faith in Big Data Must End*.

**How is this different than or similar to the problems mentioned in “The era of blind faith in big data must end”? Pick one problematic algorithm mentioned in the talk and compare its potential ethical and predictive problems to the gaydar machine.**

As Dr. O’Neil states in her TED talk, there are two parts to an algorithm: data and a definition of success. For instance, in the example of Fox News, the data were all of the applicants for jobs at Fox News from 1996 to 2017 and the definition of success was anyone who stayed at the company for four or more years and was promoted at least once. Consequently, this creates a selection bias, in that the successful people at Fox tend to be male. As mentioned above, there is a selection bias with choosing a 35,000 images subset of the 300,000 images that were originally scraped. By choosing to focus on only white, American men, the ‘Gaydar’ algorithm was built with implicit bias because it only focused on white male subset of the population.

**Find a recent article about an algorithm thought to be biased.**

This [article](https://www.computerworld.com/article/3068622/amazon-prime-and-the-racist-algorithms.html) outlines a form of modern-day redlining in which Amazon used an algorithm to deny certain zip-codes, which happened to be predominantly people of color, access to Amazon Prime’s free same day delivery. Although the Amazon spokesperson would said that the decisions to bar service to these areas was “nothing to do with race and everything to do with algorithms” (Gralla, 2016). ‘Blackbox’ algorithms of this nature are the algorithms of the cautionary tales told by Dr. O’Neil in her talk The Era of Blind Faith in Big Data Must End.

Resources:

Gralla, P. *Amazon Prime and the racist algorithms*. ComputerWorld. 2016. Accessed via: <https://www.computerworld.com/article/3068622/amazon-prime-and-the-racist-algorithms.html>

Murphy, H. *Why Stanford Researchers Tried to Create a ‘Gaydar’ Machine*. The New York Times. 2017