Brendan Jobus

18321740

One of the areas that can perpetuate inequality and bias the most is artificial inteligence, the field has grown to be one of the largest in computer science and is primarily focused on creating artificial intelligences based on things like machine learning. The way that artificial intelligences works is by creating a model that is designed to be trained by some training data, which is a set of outputs and corresponding inputs that we can use to try to get said output. These models are immensely powerful, and are used in things like image recognition to automate things that are far to tedious and infeasible to have someone do by hand. However, this leaves one big problem, if the data that we teach these models includes some inherent biases, then our model will perpetuate these these biases. As we automate systems more and more, we could see things like loan grants being automated, with aritificial intelligences being one of the best ways to automate these processes, and if the data that we feed it is biased, either with a lack of data on a minority group, or by the data itself being biased against them, then we could see minority groups struggling to get loans from banks[1]. Without enough data on a minority group, an artificial intelligence, just like a human without a lot of information on topic, might make poor and misguided decisions. This isn’t the only way our data can be biased though, certain charateristics can essentially be substitutes for race, for example wealth and address. This is because someones address can, in many cases, identify their race more so than identifying any other charateristic of an individual, because minority groups usually live in communities primarily populated by that minority. Wealth can also effectively identify an indiviuals race as the average wealth of a white family is ten times that of a black family[2]. Because of this, when we allow an A.I. to be trained on a data set with these parameters, which are often strong predictive powers, we can end up putting a person at a disadvantage, either because they unfairly group certain peoples together, or by overly weighting values that are out of their control, and thus, cause a feedback loop like in the case of wealth as a parameter. We can also find real world examples of some of these, for example, a famous case of this would be when facebooks A.I., which is used to show you videos similar to ones you have watched, flagged a video of a black man as a video about primates, and would prompt users to watch more videos about primates[3]. This error was found to be caused by a lack of data of people of color being included in training the A.I., making it less likely to be able to correctly identify them, and causing unkown behaviour to occur.

References:

<https://credera.com/insights/racial-bias-in-machine-learning-and-artificial-intelligence>[1]

<https://theconversation.com/amp/artificial-intelligence-can-deepen-social-inequality-here-are-5-ways-to-prevent-this-152226>[2]

[https://nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html](https://www.nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html)[3]