Algorithm Bias and How to Combat it

The article, Making Algorithms Less Biased, found in the ACM TechNews explained the growing issue with data algorithms and their inherent bias. It explained how there has been a tool kit developed by GovX, the Center for Government Excellence, to help address the bias and outlines ways to combat it. The tool kit is split into two parts, the first is assessing the risk, and the second is how to manage the risk. The first part contains a series of questions split up into 4 subcategories: Impact, Appropriate Use, Accountability, and Bias. The user would assess the risk in each of these categories, and then depending on those answers, refer to the second part of the tool kit to see what steps can be taken to fix the bias. The second part has 20 mitigations that outline ways to reverse the bias.

After reading this article, I conducted additional research about other changes being made in the industry, and to what extent bias in algorithms affect technology. The results were surprising to me - I had no idea how extensive bias was in both government used algorithms and other algorithms. Bias is almost impossible to avoid in complex algorithms, since algorithms learn from the data it is presented with, and almost all data is slightly biased, or not completely representative. The algorithm learns from the data and propagates the bias. For example, there was an AI chat bot that Microsoft released on Twitter, TayTweets, and it became extremely racist and sexist in a short amount of time. In addition, many companies use recruiting algorithms to help sort through applicants, and there was a study that showed that a recruiting algorithm learned to favor white males. It eventually showed recruiters primarily white males, regardless of the experience or qualifications of the other females or applicants of different ethnicities. Another study showed bias in algorithms used in the judicial system to determine the sentencing for criminals. It found that African American criminals were twice as likely to be marked as liable to re-commit, as shown in the table below.

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

Then, I researched how to combat algorithm bias at the source. I found that the biggest problem with algorithms is that they aren't being used to solve the problem they were intended for. For example, with the racist judicial algorithm, it was originally intended to just be used to set parole conditions for inmates getting released, however, it was in fact used to sentence criminals. This discrepancy between the intended use and actual use may have been a factor for its racist tendencies. Another thing that's important in preventing bias is preventing black boxing. Black boxing is the phenomena that occurs when no one, not even the programmers of the algorithm, can follow the steps of the algorithm, and fully understand why it made the decision that it did. This is a huge problem and can lead to faulty algorithms being used and never being fixed. Another important step is to promote diversity on teams producing AI. Not just in terms of gender and ethnicity, but also in terms of roles served on the team. For example, it would benefit the technology greatly if a sociologist was a part of the team and could analyze the potential societal impact of the algorithm and possibly spot biases before they occur. Lastly, filter bubbles are an important thing to be aware of. Filter bubbles occur when the user of an algorithm interacts with only a certain subset of the results of an algorithm, so the algorithm adjusts and starts only delivering results that match that subset. The example of the biased recruiting algorithm is an example of that. Once the recruiter keeps interacting with the top results he or she is being given (white males), the algorithm will think that that is the only thing that the recruiter wants to see and will make the bias worse.

This article and the subsequent research really opened my eyes about bias in algorithms and how important it is for developers to be aware of bias and be active in combating it. In addition, the tool kit that was developed is a step in the right direction for the industry and is one of the many steps being taken to get rid of this problem. There are many research institutes that have been developed with this mission and are conducting research to have a better understanding of the impact of algorithms on society. I'm glad that I am now aware of this problem, and I am intrigued to see the future of AI and the role of bias in it.

Works Cited

Nyczepir, Dave. “Making Algorithms Less Biased.” *Nextgov.com*, Nextgov, 18 Sept. 2018, www.nextgov.com/analytics-data/2018/09/making-algorithms-less-biased/151350/.

“About.” *AI Now Institute*, ainowinstitute.org/about.html.

Baer, Tobias, and Vishnu Kamalnath. “Controlling Machine-Learning Algorithms and Their Biases.” *McKinsey & Company*, www.mckinsey.com/business-functions/risk/our-insights/controlling-machine-learning-algorithms-and-their-biases.

Angwin, Julia, et al. “Machine Bias.” *ProPublica*, ProPublica, www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Clendaniel, Morgan. “Now Is The Time To Act To End Bias In AI.” *Fast Company*, Fast Company, 23 Apr. 2018, www.fastcompany.com/40536485/now-is-the-time-to-act-to-stop-bias-in-ai.

“Could New York City's AI Transparency Bill Be a Model for the Country?” *Government Technology: State & Local Government News Articles*, Emergency Management, www.govtech.com/policy/Could-New-York-Citys-AI-Transparency-Bill-Be-a-Model-for-the-Country.html

Vincent, James. “Twitter Taught Microsoft's Friendly AI Chatbot to Be a Racist Asshole in Less than a Day.” *The Verge*, The Verge, 24 Mar. 2016, www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.

Council, Young Entrepreneur. “How Can We Eliminate Bias In Our Algorithms?” *Forbes*, Forbes Magazine, 28 June 2018, www.forbes.com/sites/theyec/2018/06/27/how-can-we-eliminate-bias-in-our-algorithms/#1e031bcd337e.