**A discussion of biased algorithms**

**Introduction**

Algorithms are specific rules and instructions used to achieve a certain purpose. Historically, they have been used to perform basic tasks efficiently, through computer programs and embedded systems. As the technology has advanced, they have been used to perform more sophisticated tasks, such as scheduling flights or inventory planning, revolutionizing many industries. In the context of today’s data and technology, the applications of algorithms are so extensive that they have become a ubiquitous part of our life as most of us interact with algorithms, daily, knowingly or unknowingly, while doing activities such as streaming videos and shopping online, googling, or sending an online application for a job or university admission.

Scientists have long ago discovered machine learning algorithms, which are used to train a computer to learn the hidden patterns in the data and perform predictive and automated reasoning tasks such as credit card fraud detection, risk score and profiling, biological data classifications, etc. However, the availability of big data in recent times has spurred the growth of the widespread use of such algorithms. Using these algorithms, companies are leveraging big data in many ways, ranging from product development that enhances user experiences to increasing operational efficiencies.

There are, however, many instances in which their use has produced undesirable outcomes, some of which have worked against social equity.  These cases establish the fact that algorithms can be biased and raise serious concerns about their use without safeguarding people from harm.

**Biased algorithms**

Alake (2020) defines algorithm bias as the lack of fairness that arises from the output of a computer system. This topic got much wider attention after Joy Buolamwini, a computer scientist and a famous activist, discovered facial recognition software detecting her dark-skinned face only after putting on a white mask, during her graduate research project at MIT in 2015 (Lee, 2020), (Buolamwini, 2019).

It was generally thought that the algorithms used in Artificial Intelligence (AI) produce a fair result because their decisions are based on pure mathematics logic. But in fact, some unfair decisions made by algorithms have led us to believe that algorithms manifest the same biases as humans, and in some cases, they amplify unfair decisions (Dickson, 2018), discriminating based on categorical distinctions that affect marginalized, poor and vulnerable people.

NarxCare, an analytics tool and care management platform that is used in almost every state by doctors and pharmacists to identify a patient’s risk of misusing opioids, is one such example of a predictive algorithm that is biased in that it discriminately flags, with a high overdose risk score, individuals having psychiatric disorders, women living with depression and post-traumatic stress disorder, and patients living with complex health conditions such as cancer and other chronic diseases (Szalavitz, 2021).

**How does algorithmic bias arise?**

Biases in algorithms can arise consciously or unconsciously. Algorithms introducing unconscious bias are cognitive or faulty by type.  Whereas, algorithms introducing conscious bias are manipulated logically. The relationship of those biased algorithms is represented in figure number 1 below.



Figure 1 Categorization of biased algorithms (Source: graphical representation of Cathy O’ Neil’s hierarchical layers of algorithms)

Cognitive algorithms inherit bias from the data that reflects social inequalities and that is fed to train the algorithms to perform a task. While bias from faulty algorithms could arise due the defects in data or algorithms.

The NarxCare algorithm is both cognitive and faulty because its bias is a reflection of the data that was fed for its training and that is skewed against women, who suffered from depression and post-traumatic stress disorder, and patients living with complex health conditions. These groups of people required long-term opioid medications due to their health conditions, but they were biasedly linked to the opioid disorder.

In the past, women were abused more than men, causing more women to suffer from depression and post-traumatic stress disorder – a fact of gender inequity in our society. Also, research says that opioid disorders are more common in people having psychiatric issues. Likewise, people with complex health issues had multiple visits to hospitals and were prescribed opioids from multiple pharmacies, and research says that doctor shoppers travel long distances to get opioids.

Therefore, the data is irrationally biased against the women and these categorical people. As this data was used to train the algorithm, it learned that women having such psychiatric issues and patients who traveled long distances for opioids are more likely to have opioid disorder, predicting high opioid overdose risk scores for these groups of people.

NarxCare is also faulty because it accounts for opioids prescribed for veterinary purposes in the records of pet owners, without being able to identify the opioids ordered for their pets, increasing the opioid overdose risk of owners, whose pets were or are on opioids medications. Consequently, this group of people is also a target of NarxCare, having high opioid overdose risks.

Manipulated algorithms introduce conscious bias as the algorithm is trained accordingly to predict the desired outcome. Generally, these types of algorithms are used in fraudulent activities. Volkswagen’s algorithm that showed the wrong values of nitrogen emission by its vehicle falls into this category.

**Negative consequences of biased algorithms**

There are many instances of biased algorithms that have brought devastating social consequences. Particularly the cognitive one that affects a marginalized and vulnerable group when used publicly, and oftentimes amplifies the problem of social inequity.

For example, if public resources related to healthcare were to be distributed based on the prediction of U.S. hospitals in 2019 for the additional need of health care, black communities would get inadequate resources as the prediction was made in favor of whites (Jee, 2019). Likewise, doctors and pharmacists can easily avoid prescribing much-needed opioids to women with psychiatric illnesses and people with complex health conditions after seeing their risk scores of NarxCare, even if the scores do not reflect the actual risk. This shows how the biased algorithm cruelly amplifies the problem by targeting vulnerable and marginalized people.

**How to eliminate algorithmic bias**

There are several ways, depending upon the application, we can minimize or eliminate algorithmic bias.

To minimize the racial bias, Ghili (2021) has proposed a train and mask approach, in which all the sensitive features are used during the training phase to eliminate the proxy, but mask those features by providing the same values while predicting the outcome. Lee et al. (2019) recommended to introduce diversity in the design of algorithms upfront so that the harmful discriminatory effects on certain protected groups, especially racial and ethnic minorities will be identified and potentially be avoided.  A third-party audit is another good practice and is useful in all types of algorithms as audits prompt the review of both input data and output decisions and provide insights into the algorithm’s behavior (Lee et al., 2019). Algorithm impact assessment (AIA) is a self-regulatory tool that is can be implemented to assess the potential harms the algorithm can cause and recommend the measures to ameliorate those harms (Kozlowska, 2021).

In the case of NarxCare, had regulatory entities prepared and implemented an algorithm impact assessment, it would have prevented the negative consequences that certain groups of people faced, while preserving the benefits it provides to the doctors and pharmacists.  One solution to minimize the bias against people with complex health issues could be adjusting their weights depending on the category of the patient’s health history.

**Closing Remarks**

Many solutions to minimize or eliminate algorithmic bias have now been identified. Developers and procurers of the algorithms should implement those solutions to promote the best practice of the algorithms, maximizing benefits, minimizing risks, and continuing innovation and creativity. Government should also intervene with its policy to discourage the unhealthy practice of algorithms, particularly the one that is used in fraudulent activities and that amplifies the social problems.

# **References**

Alake, R. (2020, April 27). Algotithm Bias in Artificial Intelligence Needs To be Discussed (And Addressed). *Towards Data Schience*. Retrieved 4 17, 2011, from https://towardsdatascience.com/algorithm-bias-in-artificial-intelligence-needs-to-be-discussed-and-addressed-8d369d675a70

Buolamwini, J. (2019, February 7). Artificial Intelligence Has a Problem With Gender and Racial Bias. Here's How to solve It. *Time*. Retrieved 4 17, 2022, from https://time.com/5520558/artificial-intelligence-racial-gender-bias/

Dickson, B. (2018, 03 26). What is algorithmic bias. *TechTalks*. doi:https://bdtechtalks.com/2018/03/26/racist-sexist-ai-deep-learning-algorithms/

Ghili, S. (2021, November 24). Can Bias Be Eliminated from Algorithms? Retrieved April 24, 2022, from https://insights.som.yale.edu/insights/can-bias-be-eliminated-from-algorithms

Jee, C. (2019, October 25). A Biased medical algorithm favored white people for health-care programs. Retrieved 4 24, 2022, from https://www.technologyreview.com/2019/10/25/132184/a-biased-medical-algorithm-favored-white-people-for-healthcare-programs/?msclkid=08caf100c44211ecbe9fabcec6d0c63b

Kozlowska, I. (2021, April 4). Algorithmic Impact Assessments - What Impact Do They Have. Retrieved May 1, 2022, from https://montrealethics.ai/algorithmic-impact-assessments-what-impact-do-they-have/

Lee, J. 8. (2020, February 8). When Bias Is Coded Into Our Technology. Retrieved 4 17, 2022, from https://www.npr.org/sections/codeswitch/2020/02/08/770174171/when-bias-is-coded-into-our-technology

Nicol Turner Lee, P. R. (2019, 05 22). Algorithmic bias detection and mitigation: Best practice and ploicies to reduce consumer harns. *Brookings*. Retrieved from https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

O'Neil, C. (2017, July 16). How can we stop algorithms telling lies. Retrieved 4 18, 2022, from https://www.theguardian.com/technology/2017/jul/16/how-can-we-stop-algorithms-telling-lies

Szalavitz, M. (2021, August 11). The Pain Was Unbearable. So Why Did Doctors Turn HeR Away. Retrieved from https://www.wired.com/story/opioid-drug-addiction-algorithm-chronic-pain/