

Tech Ethics Animated – Fairness and Justice

Transcript

A medical algorithm, trained on historical data, was designed to predict the amount of care patients would require in the future. The model was trained, or developed, using historical treatment data to learn what features of an individual's case predicted the amount of care they would need in the future. Patients with greater need would then be prioritized. While the goal was to prioritize patients based on health needs, researchers found that the algorithm consistently underestimated the needs of the sickest patients in a certain minority. The researchers used the amount of money spent on patients as a proxy for how much care they needed and the severity of the patient's case. The algorithm only further amplified long-standing racial disparities in healthcare.

Now, this doesn't seem fair. And while "That's not fair" is a commonly used phrase, it isn't quite nuanced.

To better understand what is fair and what is not, we can take a look at how scholars define justice.

Scholars differentiate between distributive and procedural justice.

Distributive justice is focused on the fairness of the outcomes produced—for example, if all students received the same grade in class. Another example would be if a cake was cut equally and given to everyone. Procedural justice focuses on the fairness of the rules established and the process of decision-making. For instance, a professor can be considered fair even if not all the students received an "A." The grading would be "fair" as long as the grading process is known, based on merits, and is clear and organized. If siblings were cutting a cake, giving more cake to someone on their birthday is a fair rule.

Here, we will look at three different philosophical approaches to fairness and justice, where each approach has a different answer to what it means to be fair and just. Each approach judges how we distribute "goods," or "things of value," and many times algorithms are distributing "things of value" such as healthcare, jobs, promotion, entrance into a country, etc.

We will look at one of the two principles proposed by John Rawls in his theory of justice. The difference principle requires that the position is open to all, and any differences

would not further harm the least fortunate. There are two concepts that are useful from Rawls in evaluating data analytics.

The open position suggests that the position of inequality—or the opportunity to get the item of value—is open to everyone. When cutting the cake, the option to receive the bigger slice of cake must be open to everyone; no one is systematically disadvantaged. In the medical triage case, everyone has an equal chance to be prioritized for being treated in the same way.

In addition, Rawls' theory ensures that the least fortunate are not systematically worse off with this program. When divvying up the cake, if the test of who gets the bigger slices of the cake goes to those that ran a mile previously, then someone who was physically unable to run a mile would be disadvantaged. The physical test to get the larger slice of cake disadvantages those who are less fortunate. In the medical triage case, this data analytics program is making the least fortunate worse off. There is a huge disparity based on race discrimination in the healthcare system, and this algorithm is making it worse.

On the other hand, Robert Nozick views inequality as a part of life and questions the reasoning behind fixing inequalities. In terms of justice, Nozick is mainly focused on whether the object of the value was acquired justly and if the transfer of such an object occurred justly, too. He focuses on how one acquires something and how the good is then transferred. If the kids were allowed to bid for a larger slice of cake, the 18-year-old who has a job and more money can fairly buy the cake as long as he or she has acquired the money fairly. However, if the 12-year-old stole the money used to buy the slice of cake, this would be unfair because the acquisition was unfair. While Nozick is helpful in other cases, it is not illuminating in the medical triage case. Nozick is useful when companies acquire data with deception or fraud as he would say that any AI program trained on such data would be unjust.

Michael Walzer's approach is pluralistic compared to Rawls' and Nozick's. Valid and fair distribution of goods can take place in different "spheres" of justice, such as education, politics, or healthcare. The problem he labels as tyranny occurs when an individual dominates one sphere of life and is then able to dominate another sphere solely based on their dominance in the first sphere. Imagine patients are prioritized based on their highest degree attained. Walzer would say that this practice is not only unfair but tyrannical. Those that dominated in the education sphere should not be allocated priority in healthcare merely because they attained a higher degree. His framework of justice tests if data from one sphere is included in the allocation decision that then measures the success or failure in another sphere.

Walzer's approach is essential for data analytics because data in a large data set from one sphere can be used by a program and later applied to another sphere. In the medical triage case, Walzer would claim that this algorithm was unfair, as those who held dominance in one sphere were able to dominate healthcare and harm people due to their race. Walzer would also question why someone who is a "winner" in the economic sphere should also win in the medical sphere. The use of a dominant good—such as money—to win in unrelated spheres—such as education or medicine—is tyranny, according to Walzer.

It is important to understand the definition of justice for the ethics of data analytics—and that what constitutes fairness and justice has been debated for decades. This does not stop us from attempting to find better, more fair solutions. Programs are oftentimes used to allocate resources and goods. We care how goods like bonuses, letter grades, and your healthcare are allocated. Whether that is completed by an organization, the government, or even your healthcare provider, you should understand how fairness and justice are defined in data analytics to ensure the adequate allocation of goods.

When designing technology and AI, one should consider Rawls' open position and least-fortunate theories, Nozick's emphasis on data collection and use without deception and fraud, and Walzer's problem of tyranny.

CREDITS

Animation Team

Michael Simon (Lead)

Josiah Broughton

Script

Brooke Anquillare

Faculty Advisor

Carolina Villegas-Galaviz, Ph.D.

Associated Readings

Martin, K. (2022). Fairness and Justice in Data Analytics. In [Ethics of Data and Analytics: Concepts and Cases](#). Taylor & Francis: USA, 249-253.

Nozick, R. (1974). [Anarchy, State, and Utopia](#). Basic Books.

Rawls, J. (1971). [A Theory of Justice](#). Belknap Press.

Walzer, M. (1983). [Spheres of Justice](#). Basic Books.

Johnson, C. (2019). [Racial Bias in a Medical Algorithm Favors White Patients over Sicker Black Patients](#). *The Washington Post*.

