Antonella Basso 7/30/20

Algorithmic Injustice: A Modern Reflection of Our Racist Criminal Justice System

Rapid advancements in modern technology and society's increasing dependence on intelligent systems, in addition to maximizing the efficiency and speed of our everyday tasks, is demonstrating to have questionable consequences. Particularly in the ever growing field of machine learning (ML), which allows for the automatic processing of massive amounts of data. Recent developments in algorithmic decision-making have granted us the means to make classifications and predictions about the world through data-driven, statistical computation. In processing large datasets, these ML algorithms are able to quickly "learn" and "recognize" trends and correlations within data to classify given inputs and predict results. This revolutionary tool has infiltrated almost all systems of governance, both public and private, as a way to determine the most efficient courses of action. However, given the significant power we give these predictive models to serve as a basis for ultimate decision-making, there are several ethical matters to consider. With the rise of "Big Data" and algorithmic decision-making, the growing instances of algorithmic injustice have been overshadowed by the field's practical advantages. Yet, the use of predictive algorithms, which often conflicts with matters of fairness and social justice, is an area that deserves much attention. The aims of this paper are to shed light on one such case, namely discriminatory analytics and algorithmic injustice in the criminal justice system's use of predictive policing; to identify the epistemic and normative concerns raised in this area using a constructed map of the ethical challenges of algorithms (by Mittelstadt, Allo, Taddeo, Wachter, and Floridi); and to propose a solution to this issue with an appeal to public reason, as supported by Reuben Binns.

In recent decades, predictive ML algorithms have become the basis for policing in numerous areas throughout the U.S. "Predictive policing", the epitome of algorithmic injustice, is an attempt to improve policing with the use of ML to find more efficient ways of allocating officers, such that the number of police in any given neighborhood is proportional to the expected number of crimes. Police departments nationwide have invested in such systems, which process historical crime rate data to forecast future geographic distributions of crime and disperse their officers accordingly. On surface level, these algorithms appear unproblematic, as they are constructed to be "neutral" in their methods and "blind" to sensitive attributes (such as race/ethnicity, gender, age, etc.), operating on strictly geographical statistics. However, this does not indicate that the algorithm isn't capable of generating its own form of bias, using geography as a proxy for race, given the country's segregated cities[1]. Moreover, in this domain, it is not the algorithm itself that poses a major problem for justice, but rather the data on which it is trained. It is no question that, historically speaking, marginalized communities have been unjustly overpoliced[2]. Therefore, irrespective of actual instances of crime, we can expect these areas to

yield more reported crime when there are more officers to discover them. Ultimately, using this past biased data to make future decisions about police allocation, produces more of the same biased data. That is, continuing to disproportionately allocate police produces the same biased information which when fed back into the algorithm, creates an adverse feedback loop and perpetuates discrimination and unjust policing[2]. Thus, the predictive outputs of such algorithms are not in fact a reflection of the way the world truly is. Instead, they are a reflection of the structurally racist attitudes of our criminal justice system. As supported by the collaborative work, *The Ethics of Algorithms: Mapping the Debate*, this is a prime example of the consequences associated with having "missguided evidence", a major epistemic concern in the ethics of algorithms. Such a concern, which involves the quality of evidence produced by an algorithm, illustrates that an algorithm's inevitable reproduction of the same kind of bias that was fed into it, reflects the underlying values and attitudes of the user[3]. In predictive policing, we see that outcomes (geographical predictions of high crime) are no more racially unbiased than inputs (reported crimes in overpoliced marginalized neighborhoods). The general claim here is that we ought to be deeply concerned about the ethical implications of using algorithms that render pre-existing biases within our data. A further ethical concern described in the map of ethical challenges of algorithms, that is of particular interest in this domain, is that of "unfair outcomes", which falls under the category of normative concerns. That is, unlike an epistemic concern, it regards the actions that follow from an algorithm's results. In particular, unfair outcomes is an immediate consequence of biased evidence in predictive algorithms, which often times, leads to discrimination. While bias forms part of the decision-making process itself and is a product of misguided evidence, discrimination regards the effects of a particular algorithmic decision and is a result of unfair outcomes[3]. As mentioned, despite the ability to construct algorithms that are blind to the sensitive attributes that contribute to discrimination, this neutrality does not prevent the unmonitored use of proxies for such attributes, nor does it filter the bias already embedded in the data. Racially biased outcomes in predictive policing evidently give rise to discriminatory actions, which is not only a violation of our constitutional ethical and legal principles, but a profound crime against humanity. Allocating police in a way that targets certain racial groups, in addition to being intrinsically immoral, leads to several devastating consequences, including "self-fulfilling prophecies and stigmatisation in targeted groups, undermining their autonomy and participation in society[3] ". Particularly, overpoliced communities continue to be severely neglected in many aspects of social and political life due to the way society views them and the way they've come to view themselves[2].

Discriminatory analytics and algorithmic injustice in the space of predictive policing, as demonstrated, is not a mere technical issue that the algorithm is capable of fixing itself. Rather, it is a broader social and ethical problem that calls for radical action and significant consideration of public opinion. Specifically, when algorithmic decisions have such substantial consequences, there must be some kind of algorithmic accountability to justify these decisions on grounds that are mutually and equally acceptable to all members of society. Thus, a viable solution to the

problems of algorithmic injustice in this domain can be reasoned with an appeal to public reason, which Binns stresses in his work. He offers an approach to algorithmic accountability that involves placing shared common principles (public reason) as a constraint on algorithmic decision-making power[4]. Binns suggests that issues caused by algorithmic decisions may be more or less alleviated if decision-makers can "account for their system's outputs according to epistemic and normative standards which are acceptable to all reasonable people". In the case of predictive policing, we may thus conclude that the usage of biased data which yields discrimination and social injustice can reasonably be contested on universally acknowledged grounds. That is, the public's general recognition of such actions as unconstitutional and fundamentally immoral. By appealing to widely accepted ethical standards in this way, we may then demand more strict regulations for algorithmic accountability such that algorithmic decisions could not easily be rejected. And in so doing, we may begin to dismantle unjust policing as a democracy and prevent further damage to our society.

It should now appear evident that the type of predictive models that I've addressed, despite their unquestionable potential, must be carefully assessed in ethical terms. It does not suffice to account for an algorithm's neutral decision-making process to justify its outcomes or the actions that follow. As shown to be the case with predictive policing, the entire system must be evaluated, including the data it takes as input and the means of acquiring it. The severe ethical ramifications of using historical, racially biased data to predict future crimes are an immediate result of the collective failure to do this. Thus, for the sake of honoring the nation's democratic commitments to equality and justice, it is imperative that we honor the lives of people by giving them the fair and just treatment they deserve. To accomplish this, we must therefore use our predictive tools carefully and establish standards of algorithmic accountability that are acceptable to all people, in an attempt to eradicate discrimination and social injustice, at the very least in the context of AI.

References:

1. O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Penguin Books, 2017.

2. Annette Zimmermann, Elena Di Rosa. "Technology Can't Fix Algorithmic Injustice." *Boston Review*, 28 July 2020, bostonreview.net/science-nature-politics/annette-zimmermann-elena-di-rosa-hochan-kim-technology-cant-fix-algorithmic.

3. Mittelstadt, Brent Daniel, et al. "The Ethics of Algorithms: Mapping the Debate." *Big Data & Society*, vol. 3, no. 2, 2016, p. 205395171667967., doi:10.1177/2053951716679679.

4. Binns, Reuben. "Algorithmic Accountability and Public Reason." *Philosophy & Technology*, vol. 31, no. 4, 2017, pp. 543–556., doi:10.1007/s13347-017-0263-5.