Lei Zhang

Artificial intelligence is affecting everyone, online and in the real world. Algorithms transform people's online footprints and activities, such as online habits, shopping records, and GPS location data, into various scores and predictions for people. These scores and predictions in turn shape the decisions that affect people's lives, where discrimination and inequity become a significant problem, whether people are aware of it or not.

In 1960, M.E. Maron's notion of probabilistic Indexing, allowing a computing machine to make a statistical inference and derive a relevance for each document as a measure of the probability that the document will satisfy the given request, is perhaps the basics of machine learning.

Over the times, automated decision-making systems centered on big data, machine learning, artificial intelligence and algorithms are increasingly being used, ranging from shopping recommendations, personalized content recommendations, and accurate advertising to loan evaluations, insurance evaluations, employee evaluations, crimes in judicial proceedings and risk assessment. More and more decision-making work is replaced by machines, algorithms and artificial intelligence, in a widely accepted notion that algorithms can bring objectivity to various affairs and decision-making work in human society.

However, upon reading Jenna Burrell's paper*, How the machine 'thinks': Understanding opacity in machine learning algorithms*, in which she discusses three forms of opacity: opacity due to corporate trade secrets or state secrets, opacity arising from technological illiteracy, and opacity arising from the characteristics of machine learning algorithms and the measurements required to apply them effectively, I have come to terms with the fact that in any scenario, the design of the algorithm is the subjective choice and judgment of the programmers, a career path I would pursue after graduation. Her paper inevitably leaves me to the question of whether programmers can unbiasedly write the existing legal or moral rules into the program. Algorithmic bias has thus become a problem that needs to be addressed. Problems such as opacity, inaccuracy, unfairness, and difficulty in review brought about by the coding of rules require serious consideration and research.

Therefore, when it is necessary to question the results of autonomous decision-making systems, such as hoping to challenge the rationality or fairness of algorithmic decisions in court? How to interpret algorithms and machine learning becomes a big problem. This opacity makes it difficult to understand the inner workings of algorithms, especially for those who are not IT savvy.

Algorithmic discrimination has been around for a long time. Image recognition software has made racist mistakes, such as Google's photo software that mistakenly labeled photos of black people as "gorillas". Men see more high-paying job ads than women on Google's ad service, which of course may have to do with discrimination inherent in the online advertising market, where advertisers may prefer to target specific ads to specific groups of people.

When artificial intelligence is used to evaluate candidates, it may lead to employment discrimination. In healthcare, AI can now predict the onset of symptoms months or even years before they appear. When artificial intelligence is evaluating candidates, if it can predict that the candidate will be pregnant or suffer from depression in the future, and exclude them, this will cause serious employment discrimination.

After reading Burrell's paper, I have several questions that need to be answered, and I believe it is particular important in my future analytic career that relies heavily on algorithms: First, can fairness be quantified and formalized? Can it be translated into an operational algorithm? Second, is there a risk in fairness being quantified as a computational problem? Third, if fairness is the goal of machine learning and artificial intelligence, who decides the factors of fairness? Fourth, how to make algorithms, machine learning and artificial intelligence have the concept of fairness and be aware of discrimination in data mining and processing?

Algorithms are only as good as the data they use. For example, if an individual's risk of crime is assessed by their food preference, it is bound to get absurd results. Also, data is often imperfect in many ways, allowing algorithms to inherit the biases of human decision makers. Furthermore, data may simply reflect the persistence of discrimination within the larger society.

Relying on algorithms, data mining and other technologies without careful consideration may exclude disadvantaged groups from participating in social affairs. To make matters worse, discrimination is in many cases a by-product of the algorithm, an unforeseen, unconscious property of the algorithm rather than a conscious choice by the programmer, making it even harder to identify the source of the problem or explain it.

People question automated decision-making mainly because the system typically outputs just a number, such as a credit score or a crime risk score, without providing the material and rationale behind which the decision was made. Traditionally, judges need to make sufficient reasoning and argumentation before making a decision, which is available to the public for scrutiny. However, automated decision-making systems do not operate in this way, and most people cannot understand the principles and mechanisms of their algorithms, because automated decision-making is often made in the "black box" of the algorithm, and the problem of opacity arises.

If people are dissatisfied with the actions of the government, they can file an administrative lawsuit; if they are dissatisfied with the judge's decision, they can file an appeal, and due process ensures that these decisions can be reviewed to some extent. But if people are dissatisfied with the outcome of an algorithm's decision, can the algorithm be subject to judicial review? In an age of algorithmic decision making, scrutiny of algorithms is extremely necessary.

However, two issues need to be addressed. First, if algorithms and models can be directly censored, to what extent do people need to censor? Censoring algorithms is extremely difficult for the technically illiterate. Second, how do people judge whether an algorithm is complying

with established legal policies? Third, how can algorithms be censored in the absence of transparency?

Opacity of algorithms is a common problem, as companies can claim trade secrets or private property over algorithms. Censorship of the algorithm can be difficult in this context. Furthermore, from a cost-benefit analysis point of view, decrypting the algorithm and thus making it transparent may be very costly, and may far outweigh the benefits that can be achieved. At this point, one can only attempt to censor opaque algorithms, which may not necessarily lead to a fair outcome.

For the universal need to build technical fairness rules in advance, ensure the realization of fairness through design transparency, accountability, and accuracy of rules written into code in automated decision-making systems, all these cannot be achieved by relying on programmers alone.