

Question: How can AI be regulated in a way that is fair to all people

- How is "fair defined"
- Who are the regulators and do their personal beliefs prevent fairness for all people

Artificial Intelligence is quickly taking over every aspect of the world. From banking, to healthcare, to advertising, AI algorithms are being implemented to better solve problems faced in these industries. Specifically in a lot of cases, large amounts of data are being used to predict the likelihood of certain events occurring in the future. This is extremely powerful as complex problems can be solved much more quickly and oftentimes more accurately. However, in many cases there is bias in the AI algorithms that are used to make these kinds of decisions. These biases can result from many different factors such as bias in the dataset or personal bias of those who are charged with writing the algorithms. The rapid rise of AI has prompted the question regarding how all of this should be regulated. Most people agree AI needs some sort of regulation to ensure fairness for all people but disagreement arises when discussing how fairness should be defined. In the paper "Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law" by Wachter et al., The authors argue that AI should be regulated in a way that achieves "substantive equality", which takes into account historical inequalities. By this definition, fairness in AI regulation could require algorithms to add or remove certain metrics that perpetuate inequality. Although this definition of fairness can be beneficial in some situations, it could lead to a decrease in accuracy for a model. Another perspective is taken in the paper "A Legal Framework for Artificial Intelligence Fairness Reporting" by Yap et al. In this paper, the authors argue that there are four main fairness metrics that can be used when talking about the fairness of AI algorithms. They argue that none of these metrics cannot individually ensure fairness and tradeoffs are required. The key to defining fairness is then full transparency from those creating the model to ensure open debate. Overall, it is very difficult to regulate AI to be fair to all and the only viable solution is full transparency of methodologies from entities employing AI to ensure open rhetoric and exchange of ideas.

The first paper from Wachter et al. discusses how models should take into account historical inequality when making decisions to prevent bias. It is well known that certain artificial Intelligence algorithms use data that exists because of historical inequality. Take for example an AI algorithm used by a bank to determine the default risk of a person applying for a loan. The credit score of the person applying for the loan is used as a heavily weighted metric in determining the default risk of that individual. Statistics show that African Americans and Hispanics have a lower credit score on average and this certainly results from historic equality against these groups. Therefore an AI algorithm that uses credit score as a heavily weighted metric will reject loans for people in these groups at a higher rate than others. Some would argue that this unfairness needs to be regulated because it discriminates against people in these groups. However, as acknowledged in this paper, if an AI algorithms required an equal proportion of loans by race to be granted, there would certainly be a higher default rate among people in those groups. Although the higher default rates may be a result of historical inequality, it is not the banks responsibility to take on additional risk to solve that problem. Instead, the authors argue that bias transforming metrics are necessary and would require entities to acknowledge the biases that exist within a system and then argue whether or not those biases are necessary. In the example above, credit score could be argued as a bias that is necessary

when evaluating loan default risk, but there could be other biased metrics that cause inequality and do not improve accuracy of the model. The solution to this problem would require developers to evaluate biases that exist within the data, argue what is necessary to ensure accuracy, and be fully transparent with the public to allow for discourse regarding what bias is and is not necessary within a system.

[A LEGAL FRAMEWORK FOR ARTIFICIAL INTELLIGENCE FAIRNESS REPORTING | The Cambridge Law Journal | Cambridge Core](#)

[Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law by Sandra Wachter, Brent Mittelstadt, Chris Russell :: SSRN](#)

[The Average Credit Score by Age, Race, State & Income \(2023\) \(badcredit.org\)](#)