Algorithm Bias Facts

1. Data ethics researchers use the term algorithmic bias to refer to situations in which a computer system makes decisions that impact different people (or different groups of people) in different ways. Let's break down this term by examining what its two words mean. We’ll start with “algorithmic,” which describes an algorithm.
2. Machine learning is a type of artificial intelligence (AI)
3. To consider how algorithms might impact different people in different ways, researcher Nicholas Diakopoulos, in Algorithmic Accountability: On the Investigation of Black Boxes from Columbia Journalism Review, categorized them into the following four types: Prioritization Classification Association Filtering.
4. Prioritization algorithms help you figure out where to direct your attention. These algorithms sort or rank data based on factors that the algorithm design specifies.
5. Classification algorithms help you sort information into categories. The algorithm design might include criteria that determine the categories. Or the algorithm might automatically create the categories based on common characteristics.
6. Association algorithms connect pieces of information that are typically either related or used at the same time. Such an algorithm is usually based on a large amount of existing information.
7. Filtering algorithms either remove or separate information that meets a certain criterion. These algorithms resemble classification algorithms. But instead of sorting many items into categories, they evaluate a binary criterion.
8. We usually define bias as a situation in which one group, person, or thing is treated differently than other groups, people, or things, respectively.
9. Unconscious bias occurs when we make biased decisions based on factors that we're not aware of. This isn't a big deal if, say, you're unfairly biased against oatmeal because the Quaker Oats man scared you as a child.
10. Still, in cases of algorithmic bias, two factors come up again and again: The people who developed the algorithm come from different backgrounds than those impacted by the algorithm. The algorithm was trained with biased training data.
11. Training data refers to a labeled dataset that’s used to teach an algorithm how to make decisions. The algorithm bases its decisions on the patterns that it identifies regarding how the data points match their labels.
12. Auditing system—that is, a proactive review of an algorithm to seek out bias. Two types of auditing systems exist: external audits and internal audits.
13. An external audit. That is, the researchers had neither a relationship to the companies that created the algorithms nor any insight into how they worked. Instead, they used their own data to seek out bias.
14. An internal audit. That is, the system will regularly test an algorithm for different types of bias. This testing occurs when the algorithm is both in development and in use.
15. Checklists for Avoiding Algorithmic Bias To avoid algorithmic bias in an existing system, you might want to use the following checklist: Investigate how the system works. Was it trained by using historical data? If so, think about biases that the historical data might contain. Also, seek out evidence that the developers have controlled for biased training data. Research similar systems to find out if they include examples of bias. Try auditing the system by testing it to find out if the accuracy of its results vary according to your input. Check that a process for contesting the results of the system exists. To avoid algorithmic bias in a system that’s in development, you might want to use the following checklist: Consider the types of people that your system will serve or impact. How do these people compare to those who are developing the system? Make sure that your team is aware of any differences and teach them to seek out potentially biased decisions. Carefully review your training and testing datasets to find out if they underrepresent or overrepresent any groups of people. Consistently document your work in plain language so that others in your organization can both review and understand it. As much as possible, choose modeling techniques that people can understand. Create a process for internally auditing your system and identify someone outside your development team who can test that process.