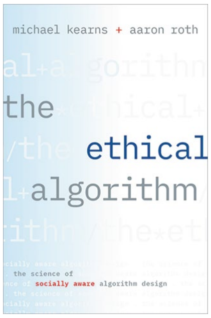
You are your own vector



The Ethical Algorithm: The Science of Socially Aware Algorithm Design

Michael Kearns and Aaron Roth

OXFORD UNIVERSITY PRESS: 2020. 256PP. £18.99

Do you side-eye Amazon's recommendations whenever you make an online purchase, or scoff at Netflix's suggestions for you? In my case, this is compounded by never-ending adverts for pregnancy tests on Youtube and fitness app ads on Instagram, regardless of the fact that I don't think I've ever clicked, rated, or purchased any of these types of products. And while targeted advertising is usually benign, I wonder how many people are as suspicious as I am about how the tech behemoths decide who we are and what we like.

This is even more relevant when much more important decisions are already being outsourced to algorithms with no human intervention — life-changing issues such as bail, parole and criminal sentencing, or more collective policies such as 'predictive policing', all of which have been shown to exhibit gender and racial bias as a feature instead of a bug.

In their book *The Ethical Algorithm*, due to be released in January 2020, Michael Kearns and Aaron Roth seamlessly weave, expand and expound on these topics and more. Two theoretical computer scientists, they are engaging and thorough as they define their aim, which is not to go over old issues with algorithms run on large datasets, but instead explain how today's algorithmic pitfalls have arisen, and what can be done about them. Rich with real-life examples ranging from famous data privacy breaches to unfair college admissions, the authors systematically break down what it is that machine-learning algorithms do, where they can go wrong and why, how we can go about fixing them, and what that will cost us.

What is most striking about *The Ethical Algorithm* is the ease with which the authors define and explain apparently nebulous concepts. They explore, for example, the many ways in which anonymized data can be breached and populate their discussion with historical examples ranging from the 1980s to today. By doing this, they make the case clear for the advantages of differential privacy. For data science aficionados, this may all be known and understood; but for someone who obsessively opts out of sending her information to Google or Apple, I will, as the French say, go to bed a wiser woman tonight. However, the authors make it clear that the power of machine learning is that it can be used to draw non-obvious correlations from the increasingly large online footprint that we generate, and that these can be very accurate predictors of aspects of ourselves that we may wish to keep separate. As the authors say: "Differential privacy doesn't protect you from these kinds of inferences, because the correlations they rely on would exist whether or not your data was available. They are just facts about the world that large-scale machine learning lets us discover."

The authors are also keen on emphasizing the difficulty in pinning down what we mean by a concept like 'fairness'. They use gender and racial bias examples to systematically cover the ways in which algorithms can produce discrimination. Even though the words "affirmative action" only appear once, the authors debunk a lot of the recent rhetoric on how, because algorithms don't use humans, they are inherently unbiased. Here, the discussion about the cost we have to pay to correct this discrimination is at its most poignant, because there is a cost: machine-learning algorithms work by minimizing error, and correcting for gender, racial, or socio-economic bias will lower accuracy. The authors deftly introduce the concept of Pareto frontiers, or a set of machine-learning models that are 'reasonable' quantitative choices for the trade-off between fairness and accuracy. While the authors are blunt in saying that fairness can be quantified just as much as accuracy, they are just as honest when they conclude that the decision on which model to use is one for society to make.

On what may seem like a lighter note, the authors use navigational apps to introduce game theory concepts (which take into account individual user preferences that may or may not be in conflict with those of others) to neatly segue into online shopping. The authors explain how Amazon sorts its users into types to produce recommendations, and easily transition into how Facebook's news algorithm turns into an echo chamber, which is not such a light note after all.

The book becomes more esoteric towards the end, where the authors explore the pitfalls of reusing existing troves of data, which may lead to false discoveries, and the morality or interpretability of algorithms that may, in the end, surpass us so that they can design themselves. While both of these topics will be an easier read for those more familiar with big data and artificial intelligence, I found them more opaque (or maybe it is just that I don't think our *The Matrix*-style overlords are coming in my lifetime).

I will say that one downside to this book is that it doesn't seem like figures are as well utilized as they could be. This stands in contrast to the writing, which is engaging and clear and could be further elevated by telling the reader where to look if there's a relevant picture.

If you are interested in your online footprint, or just confused about the adverts you receive, I cannot recommend *The Ethical Algorithm* enough. It is a thought-provoking yet easy read, demystifies the processing of large datasets and neatly lays out the power — and ultimately the limits — of designing more responsible machine-learning algorithms. ❐

Reviewed by Kristina Maria Kareh
Senior Editor, Nature Communications.
e-mail: kristina.kareh@nature.com

Published online: 8 January 2020
https://doi.org/10.1038/s41567-019-0768-1