XYZ

Lines. Words. Letters. Unlocking the News

Volume 7 | 2.10.2. 5.1.8. 9.21.1. 12.12.4. 18.12.1. 20.6.2. | Dec 13, 2024

This article contains excerpts from the New York Times "Dealing with bias in artificial intelligence" from November 2019.
Specifically the condensed comments of Dr. Koller, an adjunct professor in the computer science department at Stanford University.

You could mean bias in the sense of racial bias, gender bias. For example, you do a search for C.E.O. on Google Images, and up come 50 images of white males and one image of C.E.O. Barbie. That's one aspect of bias.

Another notion of bias, one that is highly relevant to my work, are cases in which an algorithm is latching onto something that is meaningless and could potentially give you very poor results. For example. imagine that you are trying to predict fractures from X-ray images in data from multiple hospitals. If you are not careful, the algorithm will learn to recognize which hospital generated the image. Some X-ray machines have different characteristics in the image they produce than other machines, and some hospitals have a much larger percentage of fractures than others. And so, you could actually learn to predict fractures pretty well on the data set that you were given simply by recognizing which hospital did the scan, without actually looking at the bone. The algorithm is doing something that appears to be good but it is actually doing it for the wrong reasons. The causes are the same in the sense that these are all about how the algorithm latches onto things that it shouldn't latch onto in making its prediction.

To recognize and address these situations, you have to make sure that you test the algorithm in a regime that is similar to how it will be used in the real world. So, if your machine-learning algorithm is one that is trained on the data from a given set of hospitals, and you will only use it in those same set of hospitals, then latching onto which hospital did the scan could well be a reasonable approach. It's effectively letting the algorithm incorporate prior knowledge about the patient population in different hospitals. The problem really arises if you're going to use that algorithm in the context of another hospital that wasn't in your data set to begin with. Then, you're asking the algorithm to use these biases that it learned on the hospitals that it trained on, on a hospital where the biases might be completely wrong.

Over all, there's not nearly as much sophistication as there needs to be out there for the level of rigor that we need in terms of the application of data science to real-world data, and especially biomedical data.

Reference: Smith, C.S. (2019, November 19). *Dealing with bias in artificial intelligence*. The New York Times. Retrieved July 19, 2022, from https://www.nytimes.com/2019/11/19/technology/artificial-intelligence-bias.html