NEW AMERICA WEEKLY

Edition 138

Alien Algorithms

By Cathy O'Neil

OCT. 13, 2016

Imagine a group of aliens from Mars has descended to earth. They are peaceful, nonthreatening, very smart, and have come to help us.

They hear we want to improve our system of education and tell us they have exactly the answer. They had the same problem over on Mars, they explain, and they solved it with math.

We decide to hear what they have to say. After all, there have been decades of concern over test scores, and especially over the achievement gap, which is the difference between test scores of the rich and the poor. The achievement gap is seen as a threat to our international competitiveness, and an injustice in its own right. We'd like to solve this problem efficiently and fairly, and both we and the Martians want to find a way to assess student performance equally between rich and poor districts, because the infrastructure of the whole system is biased toward the rich. Numbers, say tests scores from the same test given in all districts, should (shouldn't it?) be the best way to cut through the inherent bias. We accept the offer, and the Martian scoring systems start pushing out results. We pay close attention.

Here's how it works. Each teacher is given a score, between 0 and 100, for each class they teach. It's not entirely clear how they do it, but it's definitely better than the old system we humans had been using, where a teacher whose students didn't make some

threshold of proficiency would be blamed, or where a teacher could be fired simply by a principal's caprice. That had been obviously unfair to teachers and students, especially those in districts suffering from poverty.

And so the Martian educator assessment experts, many of whom boast impressive degrees from the most prominent Martian technical universities, hold meetings with school administrators and teacher representatives. Each teacher, they say, will be measured by how much "value" they are adding to their students. In essence, this relies on an individualized "expected score" for each student, and then a comparison of that score with the actual score achieved by that student. The teachers whose students do better than expected will be scored higher. There's lots of complicated statistical language thrown in, but the administrators trust the Martian team, which had a stellar reputation and had successfully built scoring systems on 3 other planets.

Teachers start receiving their scores by mail a few months after classes end. Some of them are surprised to learn they got terrible scores. They start trying to appeal to the Martians.

School administrators vary in how seriously they take the Martian scores. In some cases, however, they award bonuses to and fire teachers based, in part, on the Martian scores. One teacher who is fired has reason to believe that the previous year's teachers of some of her students actually cheated on those students' tests in order to get a bonus. Her evidence includes the fact that those kids, who scored quite well on their fourth grade tests, had trouble with basic reading and writing.

She tries to appeal her score but is told that, although the numbers are "suggestive" of bad numbers, she has nevertheless been treated fairly. She describes the scores like this: "I don't think anyone understands them. They have made the calculations complicated on purpose so they can't be held accountable to the validity of the results." A principal concerned about the value-added scores of her teachers asks her New York City Department of Education contact for an explanation, and is told "you wouldn't understand it, it's math."

Meanwhile, intrepid citizen journalists are starting to look into teachers' scores, even though the formulas are kept secret. Some teachers, for example, get two scores the same year because they've taught two classes—sixth grade math and seventh grade math, for example. For this population, one of four teachers gets scores that are more than 40 points apart. Indeed the scores from year to year for the same teachers are also wildly inconsistent.

The Martians maintain that it's fair, in large part because it's morally neutral. But the teachers and their advocates start formulating a response against that argument. In the aggregate, like the school or district level, these numbers might be useful to assess resource allocation. But using it to judge teachers individually is where the bureaucratic cruelty comes in. Much like infamous BMI measurement, what is useful as a collective measure can easily become misused as a personal one.

You don't actually need to imagine any of this. This is all true, except for the Martians part. The scoring system I've described above is called the "value-added model," and it's actually a class of algorithms that have become widespread in the United States over the past six years, currently used in more than half of states, often in urban districts. They were created by people who wished to fairly and objectively distinguish between successful and failing teachers.

I described it in this way not to assign blame but rather to illustrate how extreme our blind trust in algorithms has become, even when they are statistically unstable. Our faith in and fear of mathematics has allowed us replace public school accountability with secret, unfair, and unaccountable black box algorithms.

It doesn't stop at teacher assessment. We've developed all sorts of scoring systems, ranging from scores that determine who among us is likely to end up in jail, to scores that predict our response to come-ons and messages from political campaign, to scores that predict when we'll get sick, whether we'll quit our job, whether we're vulnerable to the predatory loan industry. Moreover, they've all got these three things in common: they're widespread, secret, and destructive. I have a name for algorithms like that. I call them Weapons of Math Destruction, or WMDs.

WMDs are not being developed by Martians, but by data scientists. I am myself a data scientist, so I know what the Martians do: they (we) collect some data and recognize that the best available is typically missing important features and is, more often than not, deeply biased.

Consider policing data. We know that blacks and whites smoke pot in similar numbers, but blacks are arrested twice, thrice or up to ten times more often than blacks, depending on the jurisdiction. A history of uneven policing gives rise to a racially biased digital echo in policing data. That bias is hard or impossible to untangle if you're given the job of finding patterns in the data. More often than not data scientists don't even try.

Next, you choose a "definition of success" to which to optimize your model. If you work in a company, you know that this often translates into maximizing profit, even if something that's deemed possibly profitable for your company translates into something less pleasant on the receiving end. For example, scheduling algorithms that use up-to-the-minute weather forecasts to schedule retail shift workers save money for big box stores but wreak havoc on babysitting plans. The definition of success for the model, in other words, often dooms the model to failure.

So even when they are doing their very best, data scientists can end up with an algorithm that's got a questionable definition of success and is trained by data that has cooked-in biases. This process, unsurprisingly to the careful reader of this piece but very surprisingly to those in the wider world who worship at the altar of algorithms, ends in results that may not even be meaningful, or are just a hair better than random guessing. That's fine if you're betting on the stock market with your own money, but if you're closing off job opportunities or sending people to jail, you'd want the standards to be high.

But all too often, they aren't. And they stay low in part because of the secrecy surrounding all of the above. There sometimes isn't even "ground truth" to see whether a model is doing well and how to improve it. The reasons vary, but many examples rest on the assumption that we can trust the Martians.

But we can't. So why do we want to so badly?

The more steps I take back from researching the world of algorithms, the more I recognize a pattern in the moments and situation that a WMD will deploy. It's not simply where there are people nearby who are gullible and intimidated by mathematics (that's all people everywhere in the country, probably including you). Rather, a situation is ripe for a weaponized algorithm when there's something to hide—some responsibility to be offloaded and injected into an alien black box.

That's not to say algorithms are constructed to be evil. Many of the algorithms that end up being WMDs start out life as well-intentioned plans, efforts in education or in the justice system to establish consistent, fair, and objective criteria for decisions. Algorithms are advertised as such solutions, but they don't become that way automatically. The problem is the blind faith; people are turning too much power over to the algorithm, without confirming that they are actually better than the previous system.

Algorithms form a constructed digital bureaucracy, where nobody in particular is to blame and everyone passively accepts their fate as directed from algorithmic gods on high. Seen that way, big data is a potent tool, and I expect insiders will make use of it whenever then can. Never mind that it can create systems that undermine their original goals: Instead of getting rid of bad teachers, for example, the value-added model created an atmosphere that has seen teachers flee. We are now confronting a nationwide teacher shortage.

How could we deal with this? First, we'd need to demand accountability and refuse to passively accept an opaque score. That could be done through the development of tools that allow us to interrogate the inner workings of any decision-making process, especially ones that matter a great deal to many people's lives. Critically, such auditing tools must be interpretable and relevant, the digital equivalent of the sociological experiment that tests fairness in hiring by sending resumes of similar qualifications but with black-sounding and white-sounding names.

There's a reason many of the science fiction movies end with a battle for planet Earth. After all, everybody has an agenda. The Martians may have come to help us, but their algorithms have not, not necessarily. We need to look into the black box.

Author:

Cathy O'Neil is a mathematician turned data scientist turned author. She wrote the book Weapons of Math Destruction: how big data increases inequality and threatens democracy, which was longlisted for the 2016 National Book Awards.