FUTURE TENSE THE CITIZEN'S GUIDE TO THE FUTURE.

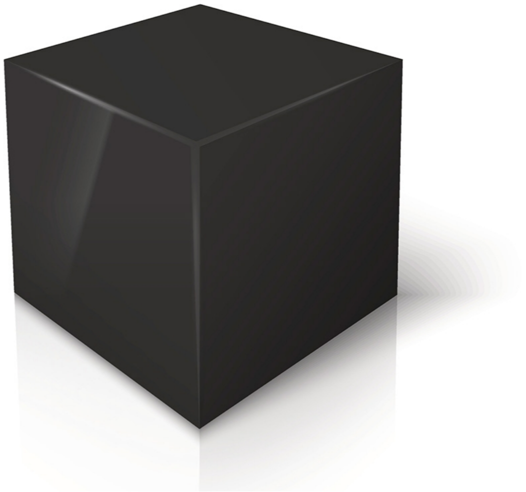
FEB. 4 2016 8:30 AM

FROM SLATE, NEW AMERICA, AND ASU

The Ethical Data Scientist

People have too much trust in numbers to be intrinsically objective.

By Cathy O'Neil



iStock

FUTUROGRAPHY

In the waning months of the Bloomberg administration, I worked for a time in a New York City Hall data group within the Health and Human Services division. One day, we were given a huge dataset on homeless families, which included various characteristics such as the number and age of children and parents, previous ZIP code, the number and lengths of previous stays in homeless services, and race. The data went back 30 years.

The goal of the project was to pair homeless families with the most appropriate services, and the first step was to build an algorithm that would predict how long a family would stay in the system given the characteristics we knew when they entered. So one of the first questions we asked was, which characteristics should we use?

Specifically, what about race? If we found that including race helped the algorithm become more accurate, we might be tempted to use it. But given that the output might decide what kind of homeless services to pair with a family, we knew we had to be extremely careful. The very fact that we were using historical data meant that we were "training our model" on data that was surely biased, given the history of racism. And since an algorithm cannot see the difference between patterns that are based on injustice and patterns that are based on traffic, choosing race as a characteristic in our model would have been unethical. Looking at old data, we might have seen that families with a black head of household was less likely to get a job, and that might have ended up meaning less job counseling for current black homeless families. In the end, we didn't include race.

Most people building algorithms (except those in politics) stay away from race for similar reasons. But what about using neighborhoods or ZIP codes? Given the level of segregation we still see in New York City neighborhoods, that's almost tantamount to using race after all. In fact most data we collect has some proxy power, and we are often unaware of it.

The question of which inputs to use in a model is actually incredible complicated, and important, given **how much we can glean about someone by their Facebook likes.** Yet it's rarely discussed. For one, people have too much trust in data to be intrinsically objective, even though it is in fact only as good as the human processes that collected it.

Another reason such ethical issues often slip by a data scientist is the power of the modern machine learning tools. Algorithms such as **neural nets** and **random forests** pick up on subtle, nonlinear connections and patterns of behavior, but in the process they lose much of their interpretability. That means nobody can exactly explain why the answer is what it is—it's almost like it's being handed down by data gods. What typically happens, especially in a "big data" situation, is that there's no careful curating of inputs. Instead, the whole kit and caboodle is thrown into an algorithm and it's trusted to come up with an accurate, albeit inexplicable, prediction.

A final obstacle to bringing up ethics in the context of data science is the training. Most data scientists are trained in applied mathematics, computer science, or statistics, fields in which an expert will publish an academic research paper that makes a tool more efficient or extends theory. Contrast that with a data scientist working inside a company or analytics group, and you see an entirely different world.

At a social media company or an online shopping site, the data scientist will run tons of experiments on users—e.g. **Facebook's famous news feed experiment** or **voter megaphone**—and depending on how they turn out, they will change the user experience to be more "optimized," however that is locally defined. A data scientist working for a local government might develop a teacher assessment that aims to sort teachers into buckets of "effectiveness," or a predictive policing model that directs law enforcement to certain areas or even certain people.

For a working data scientist, there's little time for theory. Few papers are written to describe the results of a company's experiment, even for internal use. And there is rarely ever effort put in to see how the algorithms that are deployed affect the target population beyond the narrow definition of success that the modeler is directed to track. Do new-fangled social media algorithms encourage addictive gambling behavior? Do the teacher assessments encourage good teachers to stay in education? Does predictive policing improve long-term outcomes for the people targeted by their models? Those are not the questions that a data scientist is paid to investigate.

If the data scientist's goal is to create automated processes that *affect people's lives,* then he or she should regularly consider ethics in a way that academics in computer science and statistics, generally speaking, do not. What would this look like? Let's start with low-hanging fruit: the financial crisis.

A data scientist working in finance is usually called a quant, but the job description is the same: use math, computing power, and statistics to predict (and possible to effect) outcomes. The guys (because let's be honest—they are mostly guys) working at Moody's and S&P putting AAA ratings on toxic mortgage-backed securities, for example, were data scientists. Unfortunately they were terribly dishonest; they sacrificed their integrity and scientific rigor to satisfy their bosses and to get their bonuses.

After the financial crisis, there was a short-lived moment of opportunity to accept responsibility for mistakes with the financial community. One of the more promising pushes in this direction was when **quant and writer Emanuel Derman and his colleague Paul Wilmott wrote the Modeler's Hippocratic Oath**, which nicely sums up the list of responsibilities any modeler should be aware of upon taking on the job title.

The ethical data scientist would strive to improve the world, not repeat it. That would mean deploying tools to explicitly construct fair processes. As long as our world is not perfect, and as long as data is being collected on that world, we will not be building models that are improvements on our past unless we specifically set out to do so.

At the very least it would require us to build an auditing system for algorithms. This would be not unlike the modern sociological experiment in which job applications sent to various workplaces differ only by the race of the applicant—are black job seekers unfairly turned away? That same kind of experiment can be done directly to algorithms; see the work of **Latanya Sweeney, who ran**

experiments to look into possible racist Google ad results. It can even be done transparently and repeatedly, and in this way the algorithm itself can be tested.

The ethics around algorithms is a topic that lives only partly in a technical realm, of course. A data scientist doesn't have to be an expert on the social impact of algorithms; instead, she should see herself as a facilitator of ethical conversations and a translator of the resulting ethical decisions into formal code. In other words, she wouldn't make all the ethical choices herself, but rather raise the questions with a larger and hopefully receptive group.

This issue is on the horizon, if it's not already here. The more processes we automate, the more obvious it will become that algorithms are not inherently fair and objective, and that they need human intervention. At that point the ethics of building algorithms will be taught alongside statistics and computer programming classes.

This article is part of the algorithms installment of ***Futurography****, a series in which Future Tense introduces readers to the technologies that will define tomorrow. Each month from January through June 2016, we'll choose a new technology and break it down. Read more from Futurography on algorithms:*

- **"What's the Deal With Algorithms?"**
- **"Your Algorithms Cheat Sheet"**
- **"How to Teach Yourself About Algorithms"**
- **"How to Hold Governments Accountable for the Algorithms They Use"**
- **"How Algorithms Are Changing the Way We Argue"**
- **"Which Government Algorithm to Cut Fraud Works Best—the One Targeting the Poor or the Rich?"**
- **"Algorithms Can Make Good Co-Workers"**
- **"Algorithms Aren't Like Spock—They're Like Capt. Kirk"**
- **"What Do We *Not* Want Algorithms to Do for Us?"**

Future Tense is a collaboration among ***Arizona State University****,* ***New America****, and* ***Slate****. To get the latest from Futurography in your inbox, sign up for the weekly Future Tense newsletter.*

FUTURE TENSE THE CITIZEN'S GUIDE TO THE FUTURE.

NOV. 28 2017 4:32 PM

What Happens When People Die, but Their Profiles Live On?

Readers share their most intense postmortem social media moments.

By Eleanor Cummins