Math panic

From financial crises to Facebook "filter bubbles", data scientist Cathy O'Neil sets out to expose the dangers posed by algorithms in her new book, *Weapons of Math Destruction*. Interview by **Brian Tarran**

Cathy O'Neil, a former Wall Street "quant", had a front-row seat for the financial crisis of 2007–8. She was far enough removed from the blast zone not to get vaporised in the explosion of the subprime mortgage market. But she could watch closely as the shockwaves rippled out. As wave after wave rocked the world economy, it became clear to her that mathematics could be dangerous, and algorithms destructive.

O'Neil recounts the "journey of disillusionment" that followed this experience in her new book, *Weapons of Math Destruction*, in which she argues that decision-making algorithms, fed by vast quantities of data, are increasing inequality and threatening democracy.

As a hedge fund employee, she saw how mortgage-backed securities were packaged and sold; how flawed models, built around false assumptions, were assigning triple-A ratings to "obvious garbage" and acting as a smokescreen. "I was grossed out by that," she says. "I'm a mathematician, so I like mathematics to be used fairly, and as a way of clarifying rather than obfuscating."

This was O'Neil's first interaction with a WMD – the weapons of math destruction of her book title (see box) – but it would not be her last. After leaving the hedge fund in 2009, she went to work for a risk analysis firm "with the conviction that I would work to fix the financial WMDs". Instead, she was confronted with another dangerous algorithm, and the dawning realisation that the financial services industry – despite its near-death experience – had no interest in seriously acknowledging and addressing the risks it faced.

Rebranding herself as a data scientist, O'Neil joined an online advertising company. Her job, as she describes it, involved picking "winners and losers"; she was designing algorithms to predict how likely it was that a web visitor would spend money, and those that were likely to spend would get offers that others did not. "I was giving winners the opportunities that the losers did not get, but I thought it was relatively benign," she says.

It was here that O'Neil reached the end of her journey of disillusionment, and started down a different path. "The moment that inspired me to quit my job and write this book was when this venture capitalist came to my company and was talking about the beautiful future he saw for tailored advertising … where he would get offers of trips to Aruba and jet-skis, and he would never again have to see another ad for the University of Phoenix (a for-profit university) because that wasn't for people like him."

For O'Neil, this was a wake-up call. She realised, she says, that "far from being the democratising force that we all wanted, the internet's actual purpose is to segregate us and to prey upon the poor".

But it is not just the internet, or online advertising algorithms, that O'Neil perceives as a problem. In her book, she makes the case that algorithms in many areas of life – from education, to insurance, to criminal justice – are sorting people into different "buckets" and, in doing so, rewarding certain groups while penalising others.

Paradoxically, much of this discrimination is being done in the interests of greater

O'Neil argues that decision-making algorithms, fed by vast quantities of data, are increasing inequality and threatening democracy

fairness; advocates will argue that decisions based on data and algorithms are unbiased, unlike human-led processes. But to O'Neil, that line of thinking overlooks the fact that the data we choose to collect is influenced by our own subjective opinions of what is important, and that the algorithms we design are encoded with our own agendas, beliefs and expectations.

Repeating past mistakes

Big Data is a big part of the problem, O'Neil says – particularly the idea that more data is always better. "We have all this data and we have patterns, and the machine learning algorithms are very good at pattern matching – but what they do is propagate historical practices," she says. "When we use a machine-learning algorithm to understand and 'predict' future decisions for us to make, what they are really telling us is, 'In the past, you made decisions like this based on such information'. If it's a hiring algorithm, then it might say that 'In the past, women did not do well at this company, so therefore we will filter out applications from women'." The result, of course, is that if women are not given the opportunity to succeed, the algorithm will continue to reinforce this view.

Feedback loops of this sort are pernicious, O'Neil says. Another example relates to predictive policing models (the subject of our October 2016 cover story). O'Neil describes how nuisance crimes, which are "endemic to many impoverished neighbourhoods", are fed into predictive policing algorithms, which then direct more police back into impoverished neighbourhoods. Once there, they are more likely to arrest more people for nuisance crimes, which sends more police back into those same neighbourhoods. She writes: "The result is that we criminalize poverty, believing all the while that our tools are not only scientific but fair."

To some, that might sound like a harsh judgement. After all, it can be argued that predictive policing models, like all models, are merely working with the data they have been given. But here O'Neil offers a thought experiment: "Imagine if, after the 2008 financial crisis, the police all went to Wall Street to arrest the bankers. The predictive policing algorithms would have told the police to go back to Wall Street to look for crime, as that's where it's all happening. And they would find it. My experience working on the Street is that there was more hard drug use there than I've found in other places that I've lived."

O'Neil's point, as explained in the book, is that "police make choices about where they direct their attention. Today, they focus almost exclusively on the poor. That's their heritage, and their mission, as they understand it. And now data scientists are stitching this status quo of the social order into models … that hold ever-greater sway over our lives."

The objective function

Data scientists must acknowledge their role in all this, says O'Neil. "They think of themselves as just following the numbers, but they are following the numbers to optimise their objective function," she says. Facebook, for example, wants people to stay on Facebook. That is its objective function, so its algorithm is designed to learn which friends are most important to its users, what sorts of posts users like and what articles they typically read in order to build up a profile of users so that it can present them with more stuff that is aligned with their interests.

But this feedback loop creates so-called "filter bubbles", within which users only see what they like and can only like what they see. This has implications for democracy, O'Neil says, especially when you consider how much political advertising is now directed to social media platforms. Online profiles are detailed enough that political candidates can tailor their advertising and target it in such a way that different groups of voters will be presented with messages that resonate with them, while having no idea what others are seeing. This sort of "information asymmetry … degrades our understanding of candidates' platforms", says O'Neil.

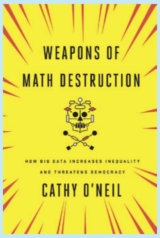
Her argument, therefore, is that a trade-off needs to be made. "We have to sacrifice accuracy for fairness sometimes," she says. More data might lead to more accurate campaign targeting, but is it fair that campaign messaging can be controlled to such an extent that voters only hear what politicians want them to hear, and that what they hear depends on the "bucket" into which they fit?

In pursuit of fairness, big data advocates must also abandon the belief that more data is always better. Here, O'Neil gives the example of orchestras, which began using blind auditions in order to eliminate nepotism in their hiring practices. Musicians were asked to perform from behind a screen, which removed all extraneous data from the hiring decision other than what was most important: the sound of the instrument. "What happened is that they got rid of nepotism but they also increased the number of women in orchestras," says O'Neil – a victory for fairness that may not have been possible had they adhered to the belief that more data is always better.

Beyond that, O'Neil believes that all of us – citizens and data scientists alike – need to get better at interrogating the algorithms we have come to rely on, and to understand that "the way we design them will inform their influence on the world". ■

What is a weapon of math destruction?

"I define a weapon of math destruction pretty precisely," says Cathy O'Neil. "They are characterised by three properties. The first is that they are widespread, they affect a lot of people in important ways – like getting a job or going to jail, or getting insurance or getting a loan, or learning about political information. The second is that they are secret, and the third is that they are destructive … they destroy people's lives unfairly, they don't solve the problem that they claim to solve, and they undermine the problem by creating a negative feedback loop."



Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, by Cathy O'Neil, is out now, published by Allen Lane