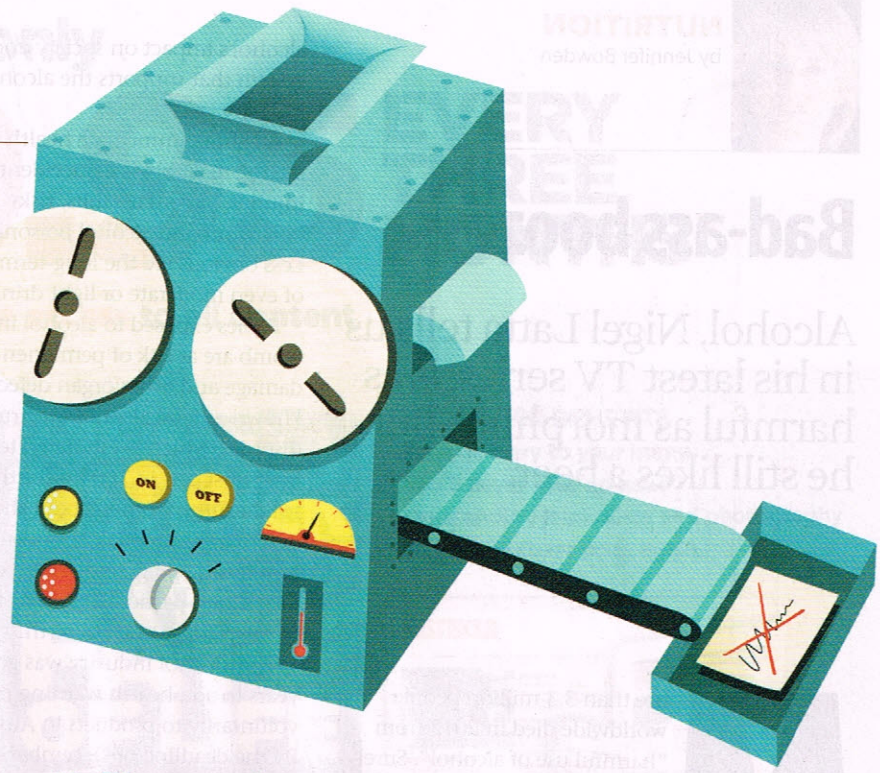
STATISTICS
by Thomas Lumley

Computer says no

When a non-driver is told he's a good car-insurance risk, you know an algorithm somewhere has stuffed up.



Modern statistics and computer science have created powerful tools for black-box prediction and classification on large sets of data. These are used for everything from spam filtering and heart attack prediction to credit risk assessment.

Unlike humans, the algorithms are blind to the meaning of the data. They have no human prejudices and just want to predict as accurately as possible.

That's not necessarily a good thing. One of the older spam-filtering tools, SpamAssassin, has an interesting warning in its documentation. If you train your spam filter on your current good email and an older database of spam email, the filter will learn that the date is the best way to detect spam and will ignore the less-reliable clues such as offers of huge inheritances or interesting pharmaceuticals. The black-box classifier just tries to match your classification as accurately as possible; it doesn't care how.

In the same way, an algorithm trained on your company's past job applicant data might notice that applicants called "John" have historically been more likely to be appointed, and later promoted, than those called "Jane", "Hone", or "Jian". The computer doesn't know that it shouldn't use this information; it just knows that's what matches your company's past behaviour.

This isn't an imaginary problem; it was encountered as early as the 1980s. Back then, St George's Hospital Medical School in London used a computer program to do initial screening of student applications to reduce staff workload. According to an editorial in the *British Medical Journal*, the program, after years of tuning, gave 90-95% agreement with selection-panel screening and it then replaced human screening.

Unlike people, computer programs aren't embarrassed by their prejudices and won't try to hide them.

Unfortunately, the program used information such as name and place of birth to screen out applicants of non-European background, as well as rejecting women. As the editors noted, this wasn't deliberate and St George's actually had a more representative student body than other University of London hospitals, but embedding the biases in software made them invisible and unchangeable.

Transparency of prediction rules is also important in situations where the appropriateness of using particular information should be open for discussion, not just when it is clearly wrong. When I lived in the US, I would get letters about once a week saying I had been identified as a low-risk driver and could get a discount on my car insurance.

I didn't have car insurance. I didn't have a car. I didn't have a driver's licence. What I did have was a good credit rating.

The insurance companies had noticed that people with good credit ratings were also, on average, less likely to have car accidents. Some people would think it's reasonable for credit ratings to be used this way, if they are informative; others would not. Without knowing what goes into a prediction, the debate can't happen.

Fortunately, it is possible to audit automated predictions much more reliably than human predictions. Unlike people, computer programs aren't embarrassed by their prejudices and won't try to hide them.

You can feed carefully selected and edited job applications or insurance forms to the algorithm to see what outputs it gives for each input. You don't need to look at how the predictions are made, so prediction rules can be evaluated for bias according to open and public criteria, even if their inner workings are a trade secret.

To find out more about the social context of Big Data, ask your favourite automated search algorithm for Cathy O'Neill, danah boyd, Ed Felten and Kate Crawford. ▮

GETTY IMAGES