

# THIS WEEK

## EDITORIALS

**TRIALS** US moves to force greater release of clinical results **p.450**

**WORLD VIEW** Take your vicar to the lab to build understanding **p.451**



**DEATH RATTLE** Lost genes can explain divergent venoms **p.453**

## Algorithm and blues

*Powerful computer programs are helping to make decisions that affect all of our lives. To avoid bias and discrimination, greater transparency and accountability are vital.*

From time to time, scientific equations appear in the media and claim to distil the perfect way to make a cup of tea or identify the most miserable day of the year. Harmless nonsense? Not according to the critics who line up on social media and blogs to complain about the pseudoscience and the commercial interests of those often involved.

Some of that scrutiny deserves a more important target. In a short space of time, the equations of big-data algorithms have permeated almost every aspect of our lives. A massive industry has grown up to comb and combine huge data sets — documenting, for example, Internet habits — to generate profiles of individuals. These often target advertising, but also inform decisions on credit, insurance and more. They help to control the news or adverts we see, and whether we get hired or fired. They can determine whether surveillance and law-enforcement agencies flag us as likely activists or dissidents — or potential security or criminal threats.

It's not just popular scrutiny that is lacking. Largely absent from the widespread use of such algorithms are the rules and safeguards that govern almost every other aspect of life in a democracy: adequate oversight, checks and balances, appeals, due process, and the right to have past offences removed from records after a statutory time.

Algorithms, from the simplest to the most complex, follow sets of instructions or learn to accomplish a goal. In principle, they could help to make impartial analyses and decisions by reducing human biases and prejudices. But there is growing concern that they risk doing the opposite, and will replicate and exacerbate human failings (see also J. T. Wilbanks and E. J. Topol *Nature* **535**, 345–348; 2016). And in an era of powerful computers, machine learning and big data, these equations have taken on a life of their own.

### BIAS IN, BIAS OUT

In some parts of the United States, the judiciary uses services — often provided by commercial companies — that use algorithms to predict the likelihood of someone reoffending. In turn, these are used in sentencing decisions, such as on whether someone gets probation or parole. Yet the results are controversial, and critics have highlighted the risk of bias against black people (claims disputed by the company that supplies the system). Similar techniques are being adopted by agencies for state surveillance and law enforcement.

There are many sources of bias in algorithms. One is the hard-coding of rules and use of data sets that already reflect common societal spin. Put bias in and get bias out. Spurious or dubious correlations are another pitfall. A widely cited example is the way in which hiring algorithms can give a person with a longer commute time a negative score, because data suggest that long commutes correlate with high staff turnover.

This risks discrimination against poorer people, often from minorities who tend to live further from central business districts. This, in

turn, could exacerbate unemployment in these areas and generate a vicious circle. Algorithms using crime and other data are also susceptible to self-fulfilling prophecies that discriminate against poorer or minority areas. A big problem is that people usually have no way of knowing what their profiles are based on — or that they exist at all.

There is an asymmetry in algorithmic power and accountability that lawmakers should correct. At the very least, there should be broader discussion of the principle that personal data belongs to an individual.

**“A simplistic over-reliance on algorithms is heavily flawed.”**

People should have the right to see their own data, how profiles are derived and have the right to challenge them. Some researchers argue that although the Internet and social media have brought benefits to democracy, recommendation algorithms can also damage the fabric of society — for example, by giving oxygen to extreme views, and by privileging sensational and superficial news or rumours that are downright false or misleading.

As Katharine Viner, editor of *The Guardian*, pointed out in July, this is being compounded by personalization algorithms that are designed to deliver what the algorithm calculates individuals want. But this tends to reinforce pre-existing views and creates echo chambers where falsehoods and irrationality can prosper.

Fortunately, a strong movement for greater ‘algorithmic accountability’ is now under way in academia and, to their credit, parts of the tech industry such as Google and Microsoft. This has been spurred largely by the increasing pace and adoption of machine learning and other artificial-intelligence (AI) techniques. A sensible step in the direction of greater transparency would be for the designers of algorithms to make public the source of the data sets they use to train and feed them. Disclosure of the design of the algorithms themselves would open these up to scrutiny, but is almost certain to collide with companies’ desire to protect their secrets (and prevent gaming).

Researchers hope to find ways to audit for bias without revealing the algorithms. Some proposed remedies are technical, such as developing new computational techniques that better address and correct discrimination both in training data sets and in the algorithms — a sort of affirmative algorithmic action. A further research goal is how to monitor and control the behaviour of largely autonomous AI systems in which even the designers have little idea of how the machine makes decisions or reaches conclusions. That could lead to the creation of algorithms to monitor the algorithms. There is much to work on and discuss.

As with the use of science metrics in research assessment, a simplistic over-reliance on algorithms is heavily flawed. It's clear that the (vastly more complex) algorithms that help to drive the rest of the world are here to stay. Indeed, ubiquitous and even more sophisticated AI algorithms are already in view. Society needs to discuss in earnest how to rid software and machines of human bugs. ■