Computer says no: why making AIs fair, open and accountable is crucial

In the final piece of a series on artificial intelligence, Ian Sample looks for the safeguards behind the tech and finds them lacking

The Guardian Domestic edition · 6 Nov 2017

Last month, American teachers prevailed in a lawsuit with their school district over a computer program that assessed their performance. The system rated teachers in Houston by comparing their students' test scores against state averages. Those with high ratings won praise and even bonuses. Those who fared poorly faced the sack.

Some teachers felt the system marked them down without good reason, but they had no way of checking if the program was fair or faulty: the company that built the software, the SAS Institute, saw its algorithm as a trade secret.

But a federal judge ruled that use of the Educational Value Added Assessment System could violate teachers' civil rights. In settling the case, the school district paid the teachers' fees and agreed to stop using the software.

The law has treated others differently. When Wisconsin police arrested Eric Loomis in 2013 for driving a car used in a shooting, he got a hefty prison term in part because the computer algorithm Compas judged him at high risk of reoffending. Loomis challenged the sentence but his stand was rejected by the Wisconsin supreme court.

The arrival of artificial intelligence has raised concerns over computerised decisions to a new high. Powerful AIs are proliferating in society, through banks, legal firms and businesses, into the National Health Service and government. It is not their popularity that is problematic; it is whether they are fair and can be held to account.

Researchers have documented a long list of AIs that make bad decisions either because of coding mistakes or biases ingrained in the data they trained on.

Bad AIs have flagged the innocent as terrorists, sent sick patients home from hospital, lost people their jobs and car licences, had people kicked off the electoral register, and chased the wrong men for child support bills. They have discriminated on the basis of names, addresses, gender and skin colour.

Bad intentions are not needed to make bad AI. A company might use an AI to search CVs for good job applicants. If the culture at the business is healthy, the AI might well spot promising candidates, but if not it might suggest people for interview who think nothing of trampling on their colleagues.

How to make AIs fair, accountable and transparent is now one of the most crucial areas of research in the field. Most AIs are made by private firms which do not let outsiders see how they work. And many AIs employ such complex neural networks even their designers cannot explain how they arrive at answers. The decisions are delivered from a "black box" and taken on trust. That may not matter if the AI is recommending the next Game of Thrones series. But the stakes are higher if it is diagnosing illness or holding sway over a person's job or prison sentence.

Last month the AI Now Institute, at New York University, which researches the social impact of AI, urged public agencies responsible for criminal justice, healthcare, welfare and education to ban black box AIs since their decisions could not be explained. "We can't accept systems in high-stakes domains that aren't accountable to the public," said Kate Crawford, a co-founder of the institute. The report said AIs should pass pre-release trials and then be monitored so that biases and nd other faults are swiftly corrected.

Tech firms know that regulations and public pressure may soon demand AIs explain their decisions, but developers want to understand them too.

Klaus-Robert Müller, professor of machine learning at the Technical University of Berlin, trained an AI to diagnose breast cancer. To understand how their AI reached decisions, Müller and his team developed an inspection program, Layerwise Relevance Propagation. LRP was used to work out how two top-performing AIs recognised horses in a vast library of images. While one AI focused on the animal's features, the other simply used some pixels near each image – which held a copyright tag for the horse pictures. The AI worked perfectly but for spurious reasons. "It's why opening the black box is important. We have to be sure we get the right answers for the right reasons," said Müller.

In many cases the AI black box need not be opened. Sandra Wachter, a lawyer and researcher in data ethics and algorithms at the Oxford Internet Institute and Alan Turing Institute, worked with her colleagues Brent Mittelstadt and Chris Russell to develop another approach. Instead of exposing an AI's full inner workings, it figures out what it would take to change the AI's decision.

For some researchers, the time to start regulating AI has arrived. Craig Fagan, policy director at Tim BernersLee's web foundation, said: "We have seen too many slip-ups, and AI is too powerful not to have government be part of the solution."

Joanna Bryson, an AI researcher at the University of Bath, thinks AI companies might be regulated like architects, who learn to work with city planners, certification schemes and licences.

In Britain and across the continent, the general data protection regulation (GDPR) comes into force in May. It gives people the right to know when companies are making automated decisions of any importance about them; it also mentions a right to explanation and a right to challenge automated decisions.

In practice though, GDPR is far weaker than the rights suggest. The right to be informed applies before decisions are made, not after the fact. And decisions can be challenged only when they are completely automated and the outcome of the decision has legal or similarly significant effects. The obligation vanishes if there is the slightest form of human involvement. The right to explanation appears still weaker. In early drafts the European parliament proposed making the right legally binding, but it was d demoted to guideline level.

Along with Luciano Floridi and Brent Mittelstadt at the Oxford Internet Institute, Wachter has called for a European AI watchdog to police the technology. The body would send independent investigators into organisations to scrutinise their AIs. To keep people safe, AIs could be certified for use in arenas such a as medicine and criminal justice.

"We need transparency ... but above all we need a mechanism to redress w whatever goes wrong, some kind of ombudsman. It's only the government that can do that," said Floridi.