Points of View No. 2

What does it mean to be human in a machine learning age?

By Andrew Briggs

Andrew Briggs is Professor of Nanomaterials at the University of Oxford. ***Human Flourishing: Scientific insight and spiritual wisdom in uncertain times*** *by Andrew Briggs and Michael Reiss will be published by Oxford University Press in 2021.*

It has been a bad summer for the public image of algorithms. 'I am afraid your grades were almost derailed by a mutant algorithm', pupils at a school were told by the Prime Minister in August. No topic in higher education is more sensitive than who gets a place at which university, and the thought that unfair decisions might be based on an errant algorithm caused understandable consternation. That algorithms have been used for many decades with widespread acceptance for coping with examination issues ranging from individual ill health to study of the wrong set text by a whole school seems quietly to have slipped under the radar.

Algorithmic decision-making is not new. In Hebrew Deuteronomic law, if a man had sex with a woman who was engaged to be married to another man, then this was unconditionally a capital offence for the man. But for the woman it depended on the circumstances. If it occurred in a city, then she would be regarded as culpable, on the grounds that she should have screamed for help. But if it occurred in the open country, then she was presumed innocent, since however loudly she might have cried out there would have been no one to hear her. This is a kind of algorithmic justice: IF in city THEN woman guilty ELSE woman not guilty.

Artificial intelligence is undergoing a transition from classification to decision-making. Broad artificial intelligence, or artificial general intelligence (AGI), in which the machines set their own goals, is the subject of gripping movies and philosophical analysis, but is still a long way off. Narrow artificial intelligence (AI) is with us now, in

the form of machine learning. Where previously computers were programmed to *perform* a task, now they are programmed to *learn* to perform a task.

We use machine learning in my laboratory in Oxford. We undertake research on solid state devices for quantum technologies such as quantum computing. We cool a device to 1/50 of a degree above absolute zero, which is colder than anywhere in the universe that we know of outside a laboratory, and put one electron into each region, which may be only 1/1000 the diameter of a hair on your head. We then have to tune up the very delicate quantum states. Even for an experienced researcher this can take several hours. Our 'machine' has learned how to tune our quantum devices in 12 minutes.

The students in my laboratory are now very reluctant to tune devices by hand. It is as if all your life you have been washing your shirts in the bathtub with a bar of soap. It may be tedious, but it is the only way to get your shirts clean, and you do it as cheerfully as you can … until one day you acquire a washing machine, so that all you have to do is put in the shirts and some detergent, shut the door and press the switch. You come back two hours later, and your shirts are clean. You never want to go back to washing them in the bathtub with a bar of soap. And no one wants to go back to doing experiments without the machine. In my laboratory the machine decides what the next measurement will be.

Many tasks previously reserved for humans are now done by machine learning. Passport control at international airports uses machine learning for passport recognition. An experienced immigration officer who examines one passport per minute might have seen four million faces by the *end* of their career. The machines were trained on fifty million faces *before* they were put into service. No wonder they do well.

Extraordinary benefits are being seen in health care. There is now a growing number of diagnostic studies in which the machines outperform humans, for example, in screening ultrasound scans or radiographs. Which would you rather be diagnosed by? An established human radiologist, or a machine with demonstrated superior

performance? To put it another way, would you want to be diagnosed by a machine that knew less than your doctor? Answer: 'No!' Well then, would you want to be diagnosed by a doctor who knew less than the machine? That's more difficult. Perhaps the question needs to be changed. Would you prefer to be treated by a doctor without machine learning or by a doctor making wise use of machine learning?

If we want humans to be involved in decisions involving our health, how much more in decisions involving our freedom. But are humans completely reliable and consistent? A peer-reviewed study suggested that the probability of a favourable parole decision depended on whether the judges had had their lunch. The very fact that appeals are sometimes successful provides empirical evidence that law, like any other human endeavour, involves uncertainty and fallibility. When it became apparent that in the UK there was inconsistency in sentencing for similar offences, in what the press called a postcode lottery, the Sentencing Council for England and Wales was established to promote greater transparency and consistency in sentencing. The code sets out factors which judges must consider in passing sentence, and ranges of tariffs for different kinds of crimes. If you like, it is another step in algorithmic sentencing. Would you want a machine that is less consistent than a judge to pass sentence? See the sequence of questions above about a doctor.

We may consider that judicial sentencing has a special case for human involvement because it involves restricting an individual's freedom. What about democracy? How should citizens decide how to vote when given the opportunity? Voter A may prioritise public services, and she may seek to identify the party (if the choices are between well-identified parties) which will best promote education, health, law and order, and other services which she values. She may also have a concern for the poor, and favour redistributive taxation. Voter B may have different priorities, and seek simply to vote for the party which in his judgement will leave him best off. Other factors may come into play, such as the perceived trustworthiness of an individual candidate, or their ability to evoke empathy from fellow citizens.

This kind of dilemma is something machines can help with, because they are good at multi-objective optimisation. A semiconductor industry might want chips that are as small as possible, and as fast as possible, and consume as little power as possible, and are as reliable as possible, and as cheap to manufacture as possible, but these requirements are in tension with one another. Techniques are becoming available to enable machines to make optimal decisions in such situations, and they may be better at them than humans. Suppose that a machine came to know my preferences better than I can articulate them myself. The best professionals can already do this in their areas of expertise, and good friends sometimes seem to know us better than we know ourselves. Suppose also that the machine was better than me at analysing which candidate if elected would be more likely to deliver the optimal combination of my preferences. Might there be something to be said for benefitting from that guidance?

By this point you may be sucking air through your intellectual teeth. You may be increasingly alarmed about machines taking decisions that should be reserved for humans. What are the sources of such unease? One may be that, at least in deep neural networks, the decisions that machines make may be only as good as the data on which they have been trained. If a machine has learned from data in which black people have an above average rate of recidivism, then black people may be disadvantaged in parole decisions taken by the machine. But this is not an area in which humans are perfect; that is why we have hidden bias training. In the era of Black Lives Matter we scarcely need reminding that humans are not immune to prejudice.

Another source of unease may be the use to which machine learning is put for commercial and political ends. If you think that machine learning is not already being applied to you, you are probably mistaken. Almost every time you do an online search or use social media, the big data companies are harvesting your data exhaust for their own ends. Even if your phone calls and emails are secure, they still generate metadata. European legislation is better than most, but it is limited in what it protects. Targeted persuasion predates AI, as Othello's Iago knew, but machine learning has brought it to an unprecedented level of industrialisation, with some of the best minds in the world

paid some of the highest salaries in the world to maximise the user's screen time and the personalisation of commercial and political influence.

Need it be so? In some ways advances in machine learning are acting as the canary in the mine, alerting us to fundamental questions about what humans are for, and what it means to be human. The old model of *Homo economicus*—rational, selfish, greedy, lazy man—has passed its sell-by date. It is being replaced by what I like to call *Homo fidelis*—ethical, caring, generous, energetic woman and man. For as long as AGI remains science fiction, it is up to humans to determine what values the machines are to implement. If we get it right, the technologies of the machine learning age will provide new opportunities for *Homo fidelis* to promote human flourishing at its best.

Christians have been thinking about what it means to be human for two millenia, building on what came before, and so they ought to have something to contribute to how humans flourish. In *It Keeps Me Seeking*, my co-authors and I ask our readers to imagine that they were writing about three thousand years ago for people who knew nothing of modern genetics or psychological science about what it means to be human. 'You are writing for a storytelling culture, and so you would probably put it in the form of a story. Let's say you set it in a garden. The garden is pleasant, but it is also designed for character formation, and so there is work to do, and also the possibility for a hard moral choice. You want to convey that humans need social interactions (for the same reason that solitary confinement is a severe punishment), and so you try the literary thought experiment of having one solitary man and letting him encounter animals and name them. Animals can be useful and they can be good company. But ultimately no animals, not even a cat, are fully satisfactory as partners in work and companions in life. Humans need humans. An enriching component of human relationships is sex. So the supreme gift to the solitary man in our story is companionship with an equal who is both like and unlike; a woman. It is hardly a complete account, but it is a good start. Oh, and there is one other aspect. They should be free of the shame which lies at the root of so much psychological disorder.'

As far as it goes, would you regard such an account as complete? If not, what would you add next? You can see where this is going. To be human you need to be responsible. So you let the humans face the moral choice. You can even include an element of disinformation to make the choice harder. And then when it goes horribly wrong you let them discover that they are responsible for their actions, and that blaming one another does not help. If you have God in your story, then (uniquely for the humans) responsibility consists of accountability to God. This is how human distinctiveness was addressed in early Jewish thought. As an early articulation that to be human means to be responsible, the story of Adam and Eve is unsurpassed.

In *Greed is Dead*, Paul Collier and John Kay reference *Citizenship in a Networked Age* as brilliantly elucidating the issue of morally pertinent decision-taking. They write, 'Whatever the future capabilities of machines, they cannot be morally load-bearing because humans are self-aware and mortal, whereas machines are not. Machines can be used not only to complement and enhance human decision-making, but for bad: search optimisation has already morphed into influence-optimisation. We must keep morally pertinent decision-taking firmly in the domain of humanity.'

The nature of humanity includes responsibility—for wise use of machine learning and much more besides. Accountability is part of life for people with widely differing philosophical, ethical, and religious world views. If we are willing to concede that accountability follows responsibility, then we should next ask, 'Accountable to whom?'

It Keeps Me Seeking: The Invitation from Science, Philosophy and Religion, by Andrew Briggs, Hans Halvorson, and Andrew Steane, is published by Oxford University Press (2018). *Citizenship in a Networked Age*, by Dominic Burbidge, Andrew Briggs, and Michael J. Reiss, is available for downloading at https://citizenshipinanetworkedage.org/.