



Rights for Robots

Simon Chesterman*

Abstract: As computers surpass human intelligence and take on greater responsibilities, should they be given rights also?

Earlier this year, the Bank of England introduced a new [fifty pound note](#). The last to switch from paper to polymer, its design was the subject of a public consultation on the question of who should be featured. More than two hundred thousand nominations were submitted—including the inevitable 'Poundy McPoundface'—but the final [decision](#) was that Alan Turing would be the face of the new currency.

Turing was a hero for his codebreaking during the Second World War. He also helped establish the discipline of computer science, laying the foundations for what we now call artificial intelligence (AI). Among his contributions is a test for when true 'intelligence' has actually been achieved.

He modelled it on a parlour game popular in 1950. A man and a woman sit in a separate room and provide written answers to questions; the other participants have to guess who provided which answer. Turing posited that a similar '[imitation game](#)' could be played with a computer. When a machine could fool people into believing that it was human, we might properly say that it was intelligent.

* Simon Chesterman is Dean and Professor of the National University of Singapore Faculty of Law. This post draws heavily on an article that first appeared in the [Straits Times](#) on 30 July 2019. The ideas are explored at greater length in [We, the Robots? Regulating Artificial Intelligence and the Limits of the Law](#) (Cambridge University Press, 2021).

I think, therefore...

Parlour games aside, why should we care about whether a computer is 'intelligent'?

As AI systems become more sophisticated, there are at least two reasons why we might start thinking of them as intelligent—and, indeed, as persons in their own right.

The first is so that we have someone to blame when things go wrong.

Ten years from now, if an autonomous vehicle crashes into you, it may be pointless to fault the human 'driver' as the car **may not even have a steering wheel**. Law reform bodies are already suggesting the concept of an automated driving system entity (ADSE) to govern liability questions. Similarly, if an autonomous agent—a lethal autonomous weapon system, say—commits a war crime, it may be [impossible to tie criminal conduct](#) to an identifiable person or corporation. Some experts have proposed that such machines could be 'punished' by being reprogrammed—or, in extreme cases, destroyed—though this is little comfort to victims and may not be an effective deterrent.

The second reason is so that we have someone to reward when things go right. Not long ago, a novella written by AI was **shortlisted** in a Japanese literary contest. Computers are increasingly engaging in quasi-creative activities. Who should own such creations?

Consider the world's **most famous selfie**—of a black crested macaque. David Slater went to Indonesia to photograph the endangered monkeys, which were too nervous to let him take close-ups. So he set up a camera that enabled them to snap their own photos.



Macaca nigra self-portrait, [Wikimedia Commons](#)

But who owned the images: Slater? The monkeys? No one?

Slater did eventually win, reflecting existing law. But as computers generate more content independently of their human programmers, it is going to be harder and harder for us to take credit. Instead of teaching a monkey how to press a button, it will be more like a teacher trying to take credit for the work of his or her student.

We hold these truths to be self-evident...

If one believes that some kind of personhood is appropriate, how should we decide which machines get it?

The Turing Test offers one approach. If people can't tell the difference between a machine and a human, maybe that's when the machine deserves to be treated like one.

The earliest successes in the Turing Test came in the 1960s with programs like *Eliza*. Users were told that Eliza was a psychotherapist, who communicated through words typed into a computer. In fact, 'she' was a program using a simple list-processing language. If the user typed in a recognised phrase, it might be reframed as a question. So after entering 'I'm depressed,' Eliza might reply 'Why do you say that you are depressed?' If it didn't recognise the phrase, it would offer something generic, like 'Can you elaborate on that?'

Even when they were told how it worked, some users insisted that Eliza had 'understood' them. Nevertheless, today no one would argue that Eliza should be treated as a 'person', though the [rise of robot nurses](#) has [challenged current notions of patient care](#).

So how should we think about the personhood of our metal and silicon counterparts?

A major hurdle is that the more we learn about machine intelligence, the less it seems we understand our own intelligence. That is all the more true of what we call 'consciousness'

But as a legal category, we do have ways of approaching the question of personality.

There are essentially two types of person recognised by law. Natural persons—humans like you and me—are recognised because of the simple fact of being human. Juridical persons, by contrast, are non-human entities—companies, for example—that are granted certain duties and rights by law.

It might seem self-evident that a computer couldn't be a natural person. Yet, for centuries, slaves and women weren't recognised as full natural persons either. If we take the Turing Test to its natural, *Blade Runner*, conclusion, it is possible that AI systems truly indistinguishable from humans might claim the same status.

For the time being, however, we are likely to stick with the possibility of juridical personhood. Unlike natural persons, the law does not recognise juridical persons for their inherent qualities but for their instrumental value. Limited liability

corporations were created to encourage entrepreneurship. If sued for wrongdoing, the corporation itself is liable rather than its investors.

Such arguments might apply to certain forms of AI. Given the diversity of AI systems, however, it is simplistic to lump all into a single category of 'person'. Implicit in many of those arguments is an assumption that AI research is on a path towards some kind of AI consciousness and a claim to natural personhood. That will probably remain science fiction—or, at least, it is not a sensible basis for regulation today.

The better solution is to rely on existing categories, with responsibility for wrongdoing tied to users, owners, or manufacturers rather than the AI systems themselves. Driverless cars are following this path, for example, with a likely shift from insuring drivers to insuring vehicles.

As for ownership of intellectual property created by AI, the moral and economic reasons for protecting human creations don't seem to apply. Would it be unjust, or would it stop AI systems doing creative work if they 'knew' that their outputs could be used by anyone?

Machines like us

In the 1990s a prize was established to encourage more serious attempts at the Turing Test. One of the first winners succeeded in part by tricking people—the program made spelling mistakes that testers assumed must have

Though the Turing Test remains a cultural touchstone, it is far from the best measure of AI research today. As a [leading textbook](#) notes, the quest for flight succeeded when the Wright brothers stopped trying to imitate birds and started learning about aerodynamics. Aeronautical engineers today don't define the goal of their field as making

Yet the more fitting tribute may be Ian McEwan's most recent novel, *Machines Like Me*, which imagines an alternative timeline in which Turing lived and was rewarded with the career and the knighthood he deserved.



The novel takes seriously the brooding syn

The reality, of course, is that chess is not a representation of life. Life is an open system; it is messy. It is also unpredictable. In the novel, the first priority of the AI robots is to disable the kill switch that might shut them down. Yet most of them ultimately destroy

Before asking whether we could create such thinking machines, McEwan reminds us, we might want to pause and ask whether we should.