

# The robot century

Robots have been part of human culture for 100 years. How can we ensure that they support – rather than supplant – humans over the next 100?



**Simon Chesterman**

For *The Straits Times*

The word “robot” entered the modern lexicon 100 years ago today with the premiere at Prague’s National Theatre of Karel Capek’s play, *R.U.R.* Set on an island “somewhere on our planet”, Rossum’s Universal Robots recounts the creation of *roboti*. Not so much mechanical creatures as stripped-down versions of humans, they were biological entities created to be strong and intelligent, but without souls.

Though dated in many ways – the limited humour derives from six men on the island vying for the hand of the only woman – the play was prescient in its vision of a world in which automatons are entrusted with serving ever more of humanity’s needs and, eventually, fighting its wars.

Reviews of the New York production called it a “brilliant satire on our mechanised civilisation; the grimmest yet subtlest arraignment of this strange, mad thing we call the industrial society of today”.

A century later, debates over the place of robots in society still echo themes in the play: how to take advantage of the benefits of technology without unacceptable risk; what entitlements are owed to entities that at least mimic and perhaps embody human qualities; what place is left for humanity if and when we are surpassed by our creations.

## NEW RULES

For the better part of that century, rules to govern robots relied heavily on science fiction. Isaac Asimov’s three laws – don’t harm humans, obey orders, protect yourself – became a cliché, unhelpful not least because they weren’t really laws at all: They constrained what his positronic creations could do, rather than saying what they should.

As artificial intelligence (AI) made truly autonomous machines



A robot serving customers earlier this month at Hajime Robot Restaurant, a Japanese eatery in Bangkok. As robots play a greater role in society, we will need to adapt our laws to accommodate them. As long as there is transparency and we maintain human control, our legal systems will cope. PHOTO: EPA-EFE

a more realistic prospect, scientists began warning that actual rules might be required.

The past few years have seen a proliferation of guides, frameworks and principles. Some were the products of conferences or industry associations, like the Asilomar and Beijing AI Principles. Others were written by companies like Microsoft and Google.

Governments have been slow to pass laws of general application, but a few – notably Singapore, Australia and New Zealand – have developed softer norms. The European Union, the G-7 and the Organisation for Economic Cooperation and Development have texts. Even the Pope endorsed a set of principles last year in the Rome Call for AI Ethics.

Virtually all include variations on the following six themes:

- Human control – AI should augment rather than reduce human potential.
- Transparency – AI should be capable of being understood.
- Safety – AI should perform as intended and be resistant to hacking.
- Accountability – Remedies should be available when harm results.
- Non-discrimination – AI systems should be inclusive and “fair”, avoiding impermissible bias.

- Privacy – The data that powers AI, in particular personal data, should be protected.

None of this seems controversial. Yet, for all the time and effort convening workshops and retreats to draft these documents, curiously little energy has gone into implementing them.

A different question yields a more revealing answer, which is whether any of these principles are, in fact, necessary. Calls for accountability, non-discrimination and privacy amount to demands that those making or using AI systems comply with laws already in place in most jurisdictions. Safety requirements would typically be covered by product liability rules.

Transparency is not an ethical principle as such, but it’s needed if we are to understand and evaluate robot conduct. Together with human control, it could be a potential restriction on the development of AI systems above and beyond existing laws.

This much was recognised in Singapore’s Model AI Governance Framework, the second edition of which had its first anniversary last week. But what would it all mean in practice?

## TRANSPARENCY

In July 2015, a group of hackers calling themselves the Impact

Team broke into a Canadian company’s website, stealing its user database and eight years of transaction records. A few weeks later, they began posting online the personal information of more than 30 million customers.

Data breaches are not uncommon, but the company in question was Ashley Madison, whose business model was based on arranging extramarital liaisons under the slogan “Life is short. Have an affair”. (The site is blocked in Singapore.)

The details posted included not only names and billing information but sexual preferences and fantasies.

The breach was initially greeted with *schadenfreude*: A bunch of adulterers were getting what they deserved.

Yet as journalists pored over the data looking for celebrity gossip, a different news story developed. The vast majority of interactions on AshleyMadison.com were not between adulterous couples, but between humans – almost all of whom were male – and automated programs known as bots.

The following month, an unusual class action lawsuit was filed seeking compensation – not on the basis of mishandling of personal data, but of fraud. Doubling down on the US\$100 (S\$133) he had spent purchasing credits on the site to chat with “women”, Mr Christopher Russell (who had

separated from his wife when he joined the site) claimed damages in excess of US\$5 million.

These claims were settled out of court, but point to something coming to be seen as a basic right: knowing whether you are talking to a human or a robot.

That might seem to be a simple question, but AI-assisted decision-making increasingly blends human and machine. Some chatbots now start on automatic for basic queries, moving through suggested responses that are vetted by a human, escalating up to direct contact with a person for unusual or more complex issues.

Another aspect of transparency is understanding decisions when they are made. In the European Union, this has led to claims of a “right to an explanation” for adverse decisions.

That might sound appealing, but it requires you to know that you have been adversely affected – easy if you are denied a loan; harder when you are being considered for government benefits or a new job.

## HUMAN CONTROL

Human control also sounds like a good thing, but if driverless cars are (eventually) safer than human-driven cars, then it will make sense for us to let go of the steering wheel. Most people would be more wary about letting loose a

truly autonomous robotic soldier, or submitting themselves to the mercy of a robot judge.

Here it is useful to distinguish between three reasons for concern about robots and AI.

The first is managing the risks associated with new technologies like autonomous vehicles. This is ultimately a utilitarian question: how to maximise benefit and reduce harm. AI is pretty good at this kind of optimisation.

Second, however, there are some activities that we might not want undertaken by machines at all. Many governments have joined the International Committee of the Red Cross and the United Nations Secretary-General in denouncing the prospect of robots making life-and-death decisions on the battlefield as “morally repugnant”.

A third category is functions whose legitimacy depends on the identity of the actor: judges, public office-holders, and so on. Even if a robot were able to make “better” decisions, we should pause before handing over responsibility for such decisions, unless we are also prepared to transfer political control – giving up the ballot box for the Xbox.

## WE, THE PEOPLE

As robots become more lifelike and social, additional rules will be required to protect them also – initially comparable, perhaps, to animal cruelty laws. Much as we should guard against robots being weaponised, it is important that they are not victimised either.

In *R.U.R.*, they are both. The play opens with the daughter of the company president sneaking into the factory to advocate on behalf of robots as a representative of the idealistic League of Humanity (the Czech word *robota* also means “slave”); it ends with them rising up and killing all but one of their makers.

Though technological apocalypse has been a staple of science fiction since Frankenstein and Prometheus, the chances of a Terminator-style “judgment day” are remote.

But as our own robots – and dispersed AI systems without physical bodies – play a greater role in society, we will need to adapt our laws to accommodate them.

Provided we can understand their actions and hold their owner, operator or user to account – as long as there is transparency and we maintain human control – our legal systems will cope.

For the rule of law is the ultimate form of anthropocentrism: Humans are the primary subject and object of law created, interpreted and enforced by humans – made manifest in government of the people, by the people, for the people.

The move of robots from science fiction to reality is forcing us to question this assumption of our own centrality, though it is not yet time to relinquish it.

stopinion@sph.com.sg

• Simon Chesterman is dean of the National University of Singapore Faculty of Law. His book *We, The Robots? Regulating Artificial Intelligence And The Limits Of The Law* will be published later this year.