Source: *The Straits Times,* pA28
Date: 23 March 2019

ByInvitation

Should we regulate AI? Can we?

Artificial intelligence is viewed by many as the defining technology of the 21st century. But how can we ensure that its benefits outweigh the potential risks?

Simon Chesterman

For The Straits Times

This week marked a grim anniversary of sorts for artificial intelligence (AI). On a Sunday night this time last year, Ms Elaine Herzberg stepped off an ornamental median strip to cross Mill Avenue in Tempe, Arizona. It was just before 10pm and the 49-year-old homeless woman was pushing a bicycle laden with shopping bags. She had nearly made it to the other side of the four-lane road when an Uber test vehicle travelling at 70kmh collided with her from the right. Ms Herzberg was taken to hospital but died of her injuries, unwittingly earning a place in history as the first pedestrian death caused by a self-driving car.

The opportunities and the threats of technology often advance hand in hand. The term "artificial intelligence" was coined at a 1956 conference in Dartmouth College in the United States. Twelve years later, Stanley Kubrick's film 2001: A Space Odyssey offered a chilling vision of a machine empowered to override the decisions of its human counterparts, the HAL 9000's eerily calm voice explaining why a spacecraft's mission to Jupiter was more important than the lives of its crew.

Both AI and the fears associated with it evolved swiftly in subsequent decades. Though worries about the impact of new technology have accompanied many inventions, AI is unusual in that some of the starkest recent warnings have come from those most knowledgeable about the field – Mr Elon Musk, Mr Bill Gates and physicist Stephen Hawking, among others.

So how should we regulate AI? If regulation is too strict, we constrain innovation and cede ground in what is thought to be the most important technological field of the century. If regulation is too lax, however, there is a real risk that when we try to adopt laws to govern that technology it could be too late.

1. Buckle Up

Many of the concerns about AI are linked to "general" or "strong" AI, meaning the creation of a system capable of performing any intellectual task that a human could – and raising complex questions about the nature of consciousness and self-awareness in a non-biological entity.

The possibility that such an entity might put its own priorities above those of humans is not one we should ignore (see Avengers: Age of Ultron). But there are more immediate challenges raised by "narrow" AI – meaning systems that can apply cognitive functions to specific tasks typically undertaken by a human.

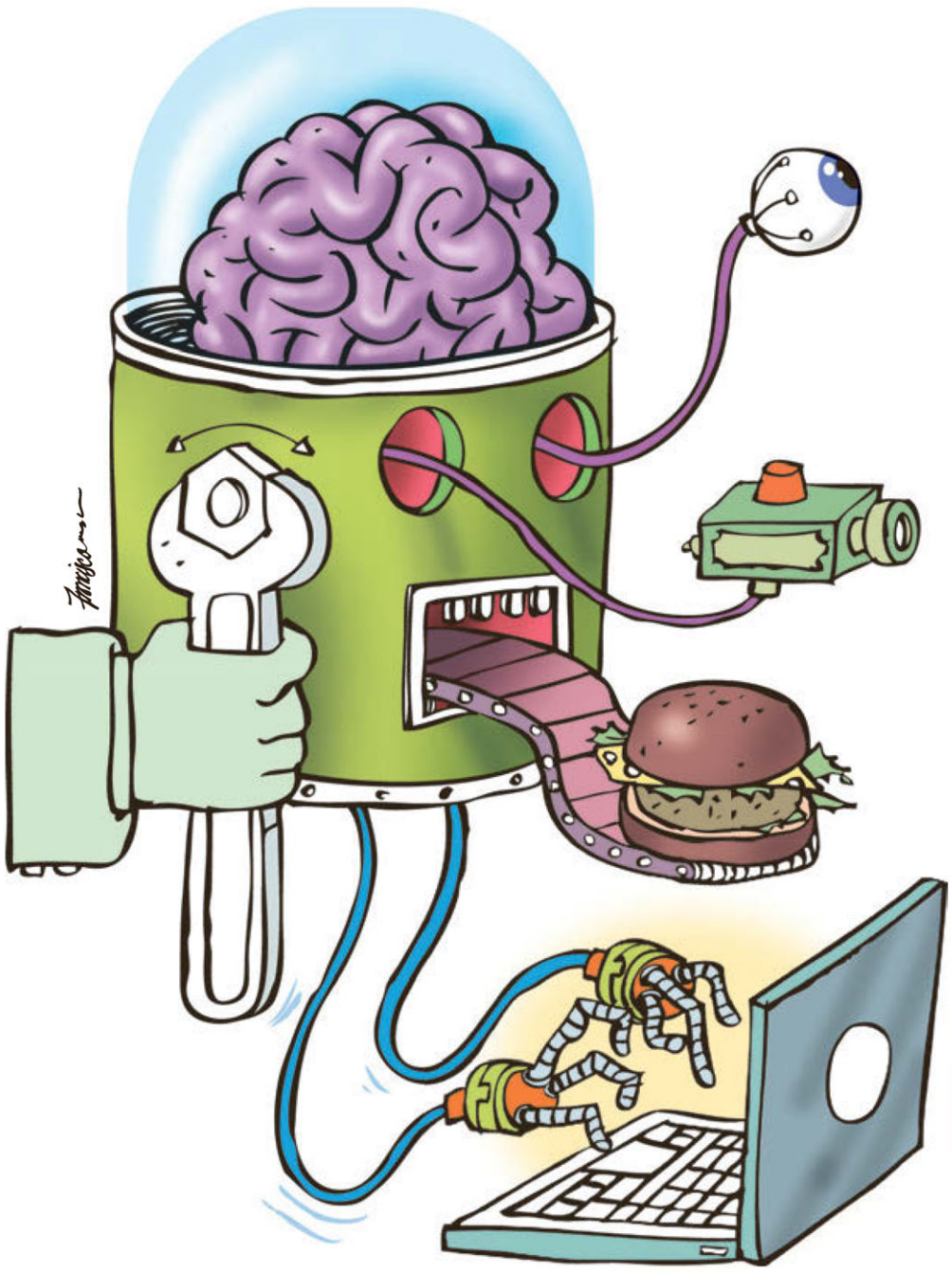
Driving, for example.

The Volvo XC90 car that struck Ms Herzberg was equipped with forward and side-facing cameras, radar and lidar (light detection and ranging), as well as navigation sensors and an integrated computing and data storage unit. A report by the US National Transportation Safety Board concluded that the vehicle detected Ms Herzberg, but that the software classified her as an unknown object, as a vehicle, and then as a bicycle with an uncertain future travel path.

At 1.3 seconds before impact, the AI system determined that emergency braking was needed – but this had been disabled to reduce the potential for "erratic vehicle behaviour".

It is still not entirely clear what went wrong on Mill Avenue that night. Uber pulled its cars from all four US cities in which they were operating, but eight months later resumed testing — though its vehicles were now limited to 40kmh and would not operate at night or in wet weather.

That lack of clarity highlights two of the main concerns about AI: the ability to operate autonomously, combined with the fact that the processes by which certain decisions are made may be opaque to outside observers and even the programmers themselves.



For this reason, a model framework for AI by Singapore's Personal Data Protection Commission (PDPC), launched by Minister for Communications and Information S. Iswaran at Davos in January, emphasises the importance of human-centricity and transparency. In essence, the individual consumer or customer should be the focus of AI design and deployment, and decisions made by AI should be explainable and fair.

2. Don't Text and Drive

The car that killed Ms Herzberg was operating autonomously, but it was not empty. Sitting in the driver's seat was Ms Rafaela Vasquez, hired by Uber as a safety driver.

The safety driver was expected to intervene and take action if necessary, though the system was not designed to alert her. Police later determined that she had most likely been watching a streaming video – an episode of the televised singing competition The Voice, it seems – for the 20 minutes prior to the crash.

System data shows that, just before impact, she did reach for the steering wheel and applied the brakes about a second later – after hitting the pedestrian. Once the car had stopped, Ms Vasquez called 911 for assistance.

Who should be held responsible for such an incident: Uber? The "driver"? The car itself? (There is also an argument that the late Ms Herzberg might have been partly at fault.)

The PDPC framework does not seek to resolve questions of legal liability, but human-centricity and transparency are useful starting points.

The law typically seeks to deter identifiable persons – including legal persons, such as companies – from certain forms of conduct, or to allocate losses when there is harm. Key questions to consider here are how the acts and omissions of AI systems can and should be attributed to traditional legal persons.

As AI systems operate with greater autonomy, the idea that they might themselves be held responsible has started to be taken more seriously. Yet this is both too simple and too complex. It is simplistic in that it lumps a wide range of technologies together in a single legal category; it is overly complex in that it assumes that AI will eventually assume full legal personality in the manner of the "robot consciousness" arguments mentioned earlier. Though not inconceivable, that is not a sound basis for regulation today or for the foreseeable future.

Notions of foreseeability underpin the other tool that has been embraced as a means of limiting the risks associated with AI: transparency. There is now a movement among developers towards explainable AI (XAI), meaning programs that can be understood and appropriately trusted by humans. In the European Union's recent General Data Protection Regulation, this has been codified as a "right to explanation". Singapore's model framework says that AI decisions should be explainable, transparent and fair.

But the limits of transparency are already beginning to show.

To pick a trivial example, when a computer beat world champion Go player Lee Sodol in 2016, its programmers could not explain how it came up with a particular strategy.

More seriously, the US Supreme Court was confronted by an appeal in 2017 against a sentencing decision based on proprietary software that analyses the risk of reoffending.

The appeal was ultimately dismissed, probably because the lower court had concluded that the decision was supported by other factors. As one of those judges stressed, a court can consider such tools – but it must not rely on them.

Anyone who has been stranded when their smartphone died knows, however, that the path to reliance on technology is a slippery one.

3. Slow Down?

Underlying the question of regulation is the need to balance precautionary steps against unnecessarily constraining innovation.

As a committee reviewing Singapore's Penal Code last August concluded, despite the risks posed by AI, it is telling that no country has introduced specific rules on criminal liability for artificial intelligence systems: "Being the global first-mover on such rules may impair Singapore's ability to attract top industry players in the field of AI."

Such concerns are well founded. Overly restrictive laws can stifle innovation or drive it elsewhere.

Yet the failure to develop appropriate legal tools risks allowing profit-motivated actors to shape large sections of the economy around their interests to the point that regulators will struggle to catch up. This has been particularly true in the field of information technology.

Social media giants like Facebook, for example, monetised users' personal data while data protection laws were still in their infancy. Similarly, Uber and other first movers in what is now termed the sharing or "gig" economy exploited platform technology before rules were in place to protect workers or maintain standards.

As the University of Washington's Professor Pedro Domingos once observed, people worry that computers will get too smart and take over the world. The real problem with computers is that they're too dumb and they've taken it over already.

4. Bumpy Road Ahead

Earlier this month, the Attorney for Yavapai County in Arizona, Ms Sheila Polk, concluded that there was no basis for criminal liability on the part of Uber in the death of Ms Herzberg. She did, however, recommend that there should be further investigation of the backup driver, Ms Vasquez, with a view to possible prosecution for manslaughter.

The Volvo XC90 itself, together with its on-board computer system, has been repaired and is, presumably, back on the road.

So, is this also the first anniversary of a machine getting away with murder? I'll let you be the judge of that.

stopinion@sph.com.sg

• Simon Chesterman is dean and professor of the National University of Singapore Faculty of Law.