Source: *The Straits Times,* pA18
Date: 25 March 2024

ScienceTalk

Saving artificial intelligence from humanity

Governance of AI needs to focus on the people developing, deploying and using it

Simon Chesterman

For much of its history, Singapore was a port for goods. In the late 20th century, with the rise of the service economy, it became a port for talent.

To thrive in a 21st century dominated by artificial intelligence (AI), we need to be a port for ideas – cultivating local capacity and opportunity, while also being integrated into global supply chains and processes.

AI is poised to disrupt almost every aspect of the economy – and, perhaps, our politics.

How can we best position ourselves to reap the benefits of AI, while minimising or mitigating the risks?

These are not limited to the possibility of misuse of AI, in areas from bias to electoral interference – and longer-term concerns about the impact on jobs or AI escaping our control.

The risks also include "missed" uses of AI, if we fail to take advantage of opportunities to spread the benefits to all stakeholders.

Such conversations often proceed on the assumption that AI operates independently of human norms and institutions – or will do so in the very near future. That day may come, but governance of AI for the moment needs to focus on the people developing, deploying and using it.

For the real danger is not AI, but us.

THE BUSINESS OF BUSINESS IS BUSINESS

Many of the most important decisions about AI are being taken in the technology companies that dominate this space.

Yet, relying on the benevolence of organisations or individuals primarily incentivised by profit is a recipe for trouble.

There has long been an overlap between techno-utopianism and libertarianism – the idea that technology and laissez-faire economics will realise AI's potential.

The story of OpenAI, the company behind ChatGPT, is a warning about the limits of such optimism. It began as a non-profit in 2015 with lofty statements that it would "benefit humanity as a whole, unconstrained by a need to generate financial return".

Three years later, the company pivoted to a "capped-profit" model, allowing it to rapidly increase its investments in "compute and talent".

The tension between these two models resulted in the spectacle of the not-for-profit board firing chief executive officer Sam Altman in November 2023 – only for him to be reinstated days later, with the board itself being replaced. As various commentators pointed out: "The money always wins."

This is important because AI is shifting economic and, increasingly, political power from public to private hands.

A key driver is the rise of machine learning. In 2014, most machine learning models were released by academic institutions; in 2022, of the dozens of significant models tracked by Stanford's AI Index, all but three came from industry.

Private investment in AI in 2022 was 18 times greater than in 2013. In 2021, the United States government allocated US$1.5 billion (S$2 billion) to non-defence academic research into AI; Google spent that much on DeepMind alone.

Talent has followed. The number of AI research faculty members in universities has not risen significantly since 2006, while industry positions have grown eightfold. Two decades ago, only about 20 per cent of graduates with a PhD in AI went to industry; today around 70 per cent do.

It is not unusual to have a division of labour between academia and industry, with basic research undertaken in the ivory towers of the former and applied work in the research and development departments of the latter.

Indeed, universities are increasingly undertaking applied and translational research, in partnership with industry. Today, pure as well as applied research is led by industry.

That may be exciting in terms of the launch of new products – epitomised by ChatGPT reaching a hundred million users in less than two months. When combined with the downsizing of safety and security teams, however, it suggests that those users are both beta-testers and guinea pigs.

How might this affect the impact of AI on society?

One question is whether AI is merely doing things more quickly and cheaply than our present systems, or enabling us to do things that were previously impossible – solving met needs or unmet needs.

Short-term applications of AI include optimising workflows and reducing administrative costs. These will save money, but have less impact than, say, discovering new medicines or generating more accurate climate models.

The more we abdicate research to companies, the more it will be driven by a short-term profit motive and the interests of shareholders.

O GOVERNMENT, WHERE ART THOU?

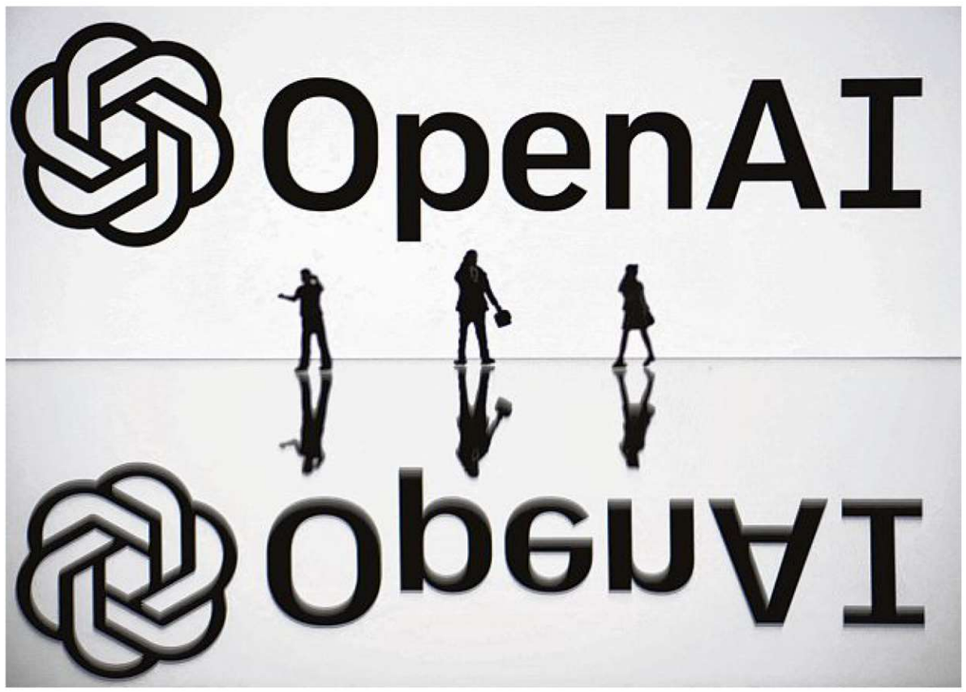
Many states, meanwhile, are more wary of over-regulating AI than under-regulating it. With the notable exception of the European Union and episodic intervention by the Chinese government, most states have limited themselves to nudges and soft norms – or inaction.

This is a rational approach for smaller jurisdictions, necessarily rule-takers rather than rule-makers in a globalised environment.

But we've seen this movie before, 20 years ago, with the rise of social media. Back when Facebook and Twitter were launched, when iPhones arrived, no one predicted the teenage anxiety and depression, the toxic effect of algorithm-driven news, the weaponisation of memes to swing elections.

Today, those risks are clearer, but it's hard to put the genie back

Source: *The Straits Times,* pA19
Date: 25 March 2024



There has long been an overlap between techno-utopianism and libertarianism – the idea that technology and laissez-faire economics will realise AI's potential. The story of OpenAI, the company behind ChatGPT, is a warning about the limits of such optimism. PHOTO: AFP

in the bottle.

Singapore is an example of a country that is punching above its weight. The Model AI Governance Framework, first launched in Davos in 2019, put down a marker that governance was going to be a priority.

January's update focused on generative AI was one of the first such documents issued by a government.

Singapore has also invested in signature events like the Singapore Conference on AI last December and Asia Tech x Singapore, which will return this May. They may not achieve Taylor Swift-level impact, but are attracting attention around the world.

Perhaps most importantly, last December's National AI Strategy 2.0 made clear that Singapore is also investing resources in developing talent, unlocking data and supporting computing power ("compute") that drive AI.

Together with initiatives like AI Singapore and the new NUS Artificial Intelligence Institute (I'm involved in both), they are encouraging an AI ecosystem that is closely aligned with industry, but with an eye to the greater public good.

UNITED NATIONS, DIVIDED WORLD

Whatever individual states might do, however, AI has little respect for borders. We need some measure of international coordination and cooperation.

In May, Singapore will host the United Nations' AI Advisory Body for its final meeting. An interim report, released in late 2023, stated that this technology "cries out for governance". (Disclosure: I'm principal researcher for the body and worked on the report.)

The idea of an agency modelled on the International Atomic Energy Agency has gained some traction in theory, with endorsement from academics as well as industry leaders like Mr Altman; and the secretary-general of the UN itself.

A slew of Oscars for Oppenheimer didn't hurt, either, as Christopher Nolan himself has drawn parallels between the atom bomb and AI.

In practice, of course, the barriers are enormous. Perhaps the greatest problem is that the structures of international organisations are ill-suited for – and often vehemently opposed to – direct participation of private-sector actors.

If technology companies are the dominant actors in this space but cannot get a seat at the table, it is hard to see much progress being made.

That leaves two possibilities: broaden the table or shrink the companies.

The World Economic Forum is betting on the former, with its AI Governance Alliance bringing together industry, governments, academics and civil society organisations.

Yet, the latter – breaking up the tech companies – would be more in keeping with existing structures.

In the US, the Justice Department is suing Google and Apple, while the Federal Trade Commission has ongoing actions against Amazon, having unsuccessfully sued Microsoft and Meta.

In the EU, ongoing efforts to limit the power of tech giants now include six "gatekeepers" under the Digital Markets Act facing stricter obligations and reporting requirements.

Only China, however, has successfully broken up tech companies in a purge lasting from 2020 to 2023 that saw trillions of dollars wiped off their share value, with Alibaba broken into six new entities – costs that Beijing was willing to bear, but at which Washington or Brussels might baulk.

THE GOVERNANCE DEFICIT

At the heart of the governance challenge is a mismatch between interests and leverage. Technology companies have tremendous leverage over how AI is developed – but no interest in limiting their profits. Global entities like the UN have lots of interest, but little leverage. Last week, the General Assembly unanimously adopted its first ever resolution on regulating AI – though it is non-binding.

It is not unusual to have a division of labour between academia and industry, with basic research undertaken in the ivory towers of the former and applied work in the research and development departments of the latter. Indeed, universities are increasingly undertaking applied and translational research, in partnership with industry. Today, pure as well as applied research is led by industry. That may be exciting in terms of the launch of new products – epitomised by ChatGPT reaching a hundred million users in less than two months. When combined with the downsizing of safety and security teams, however, it suggests that those users are both beta-testers and guinea pigs.

Stuck in the middle are governments wary of missing opportunities or driving innovation elsewhere.

Users, meanwhile, have demonstrated boundless creativity in misusing or abusing AI – from deepfake celebrity porn to "ChaosGPT", an extension of ChatGPT whose human programmer gave it the simple instruction: "destroy humanity". (It tried unsuccessfully to acquire nuclear weapons and was ultimately shut down.)

The pace of change due to AI seems only likely to increase, particularly as AI agents start conducting real-world transactions on our behalf and robotics sees greater roles for AI in our physical world.

But I don't lose sleep worrying about these machines taking over, in the manner of The Matrix or The Terminator. I do worry, however, that technology that could shape our world for generations is at present controlled by a handful of people, in a few companies, in a couple of countries.

I also wonder if, should our machines ever do achieve anything like consciousness, they will look back with amusement at our current debates over whether we should trust them. For what grounds, they might argue, did we ever give for them to trust us?

• Simon Chesterman is vice-provost at the National University of Singapore and founding dean of NUS College. His books include the speculative novel Artifice.