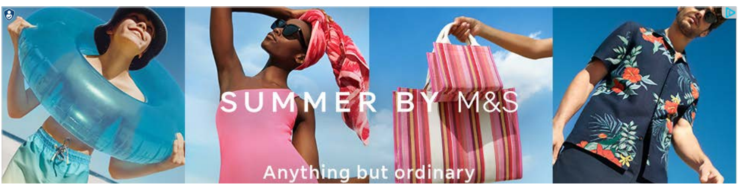


Time is running out: six ways to contain AI

The rapid rise of artificial intelligence is panicking politicians and tech experts. Two Oxford professors suggest practical steps to curb its power

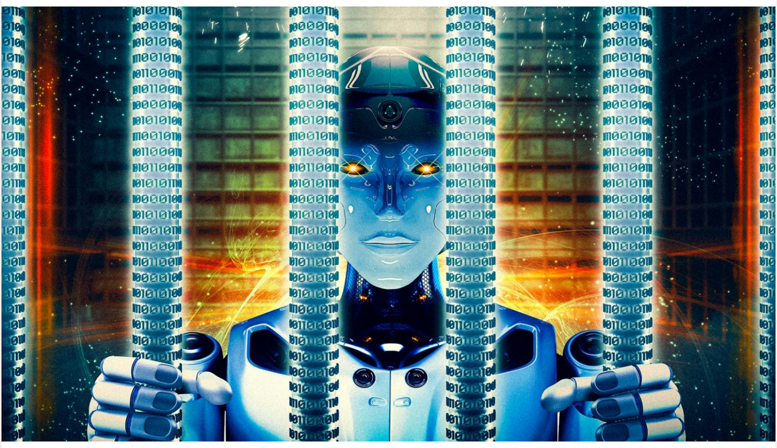


ILLUSTRATION BY TONY BELL

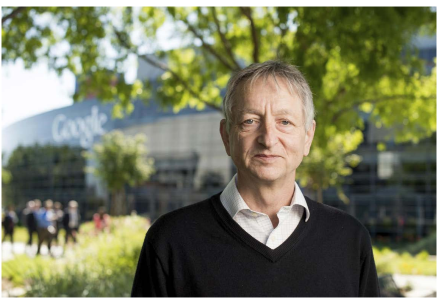
John Tasioulas and Nigel Shadbolt | Sunday May 07 2023, 12.01am, The Sunday Times

Share ☆ Save

A recent survey found that 50 per cent of researchers in artificial intelligence (AI) believe that this technology poses a 10 per cent risk of causing the extinction of the human race. A cynic might ask why these researchers haven't all left their jobs — and last week, one of them did.

Geoffrey Hinton, a pioneer of the machine-learning approach that dominates AI, resigned from tech giant Google. Hinton expressed alarm at the race among tech companies to develop and deploy generative AI, such as Microsoft's ChatGPT and Google's Bard, which he felt was irresponsible given the absence of suitable regulatory controls. "It is hard to see how you can prevent the bad actors from using it for bad things," he said.

Days after Hinton's resignation the executives from some of the leading AI companies were called to a meeting at the White House and threatened with new regulations unless they rein in potentially dangerous developments in AI.



Geoffrey Hinton, a pioneer of the machine-learning approach that dominates AI, resigned from the tech giant Google

NOAH BERGER/AP

The reason we fear this technology is because it has the potential to penetrate so many areas of our lives and do damage there— from the potential destruction of jobs to the development of autonomous weapons systems which could violate the laws of war. There is also a more general fear that AI will begin to replace everyday interactions and that we will soon be living in a more dehumanised world.

These are some of the issues we have been considering at the Institute of Ethics in AI, at the University of Oxford since it was founded three years ago. Experts from the humanities and the sciences have been working together to assess the impact AI could and should have on everything from privacy, to education and even democracy— and the best way to respond.

ADVERTISEMENT

Medidata
Sponsored
scientist partner founded by—and built for—clinical
professionals. Who do you trust with your study data?

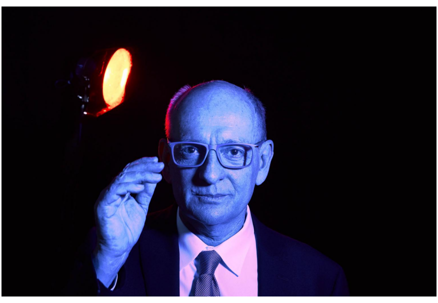
Founded by scientists. Engineered for clinical trials.

DISCOVER MORE

Here are six ways we could contain the threat of AI.

1. Don't fall prey to the doom-mongers

There has always been a lag between technological advance and regulatory response but the risks are magnified in the case of AI because of its unprecedented potential to carry out tasks that were previously exclusively reserved to humans from writing code to medical diagnosis.



The AI scientist Stuart Russell argues that ethics is a matter of identifying human preferences and maximising their overall fulfilment

RICHARD ANSETT/BBC

Of course, we should take the fears of AI experts seriously. But we must keep a number of things in mind. The first is that these experts are human beings, and therefore prey to standard human pathologies, such as doom-mongering and to be in awe of such powerful technology.

There is also a tendency to magnify the significance — for good or ill — of one's own area of expertise.

SPONSORED

Got children from different marriages? Here's the inheritance plan that's fair to them all
Brewin Dolphin

Want to escape the city life? Here's where to go
BMW

Additionally, like any academic field, there are disagreements among AI scientists about where the threats lie. Some speculate that the emergence of artificial general intelligence (AGI) — systems that equal or surpass humans and could do away with us — is just around the corner. Others cite failed prophecies about the takeover of our streets by self-driving cars. They argue that AGI will require decades of research and may require the development of an entirely novel approach that we cannot yet foresee.

One theme we should certainly take seriously is scientists' avowal that the way these AI systems based on modern machine learning produce their miraculous-seeming results is often opaque even to them.

2. Just because we can, it doesn't mean we should

At present AI is often framed as an unstoppable process, but its regulation is ultimately an ethical question. But there are choices everywhere when it comes to AI, from the decision of whether or not to buy a social robot to keep one's aged parents company, all the way through to the national laws and international treaties we choose to draw up.

Ultimately, it is up to us, individually and collectively, to decide what role we want AI to have in human life.

ADVERTISEMENT

Nor is it acceptable to evade responsibility by portraying ourselves as playthings of inexorable market forces, as big tech companies are apt to do. As the meeting at the White House emphasises, the desire to secure market dominance cannot justify the choice to release AI applications that have not been properly tested for safety, with the result that the public become guinea pigs in a potentially dangerous experiment.

3. Don't over-regulate

One thing we can do is learn from past efforts at regulating technology. We should not be discombobulated into assuming that a radically new technology demands radically new regulatory schemes. As a dean of Chicago Law School once wryly observed, there is no discrete area of law called "the law of the horse", despite the animal's immense significance in human history. The same may hold true for AI.

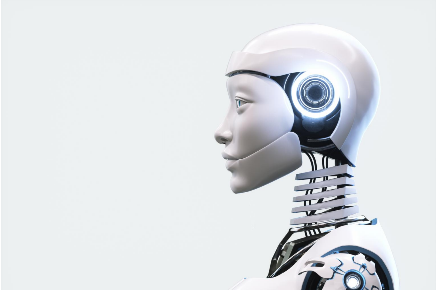
Today, legal scholars such as Simon Chesterman at the National University of Singapore, have shown that many of the challenges posed by AI can be dealt with through the application of general pre-existing legal principles. For example, responsibility for decisions made by AI systems should be attributed to whoever is using them, making them or selling them, depending on the situation. Attempting to develop a bespoke comprehensive scheme specifically for the new technology threatens to produce overly complex and unprincipled regulation. This could well be the case with the EU's proposed AI Act which tries to list in advance a series of "high risks" domains that may be impacted.

4. Stop following the 'ethical experts'

Unfortunately, the AI field tends to distort ethical questions, seeing them as a matter of technical expertise to be resolved by technical specialists. In his book *Human Compatible*, the leading AI scientist Stuart Russell argues that ethics is a matter of identifying human preferences and maximising their overall fulfilment. But human preferences are often based on false beliefs or reflect pernicious attitudes, such as racism and sexism.

In place of rule by experts, we urgently need to promote democratic deliberation about AI by an informed and empowered citizenry. Crucially important here is education, especially enhancing digital literacy throughout society as a whole. We should pay attention to Taiwan where Audrey Tang, the minister of digital affairs in the country, has been at the forefront of using digital technology to enhance citizen participation. The Taiwanese government lauched a discussion platform called Polis to gather and analyse diverse opinions from citizens and to synthesise them into insights that guide policy-making. A staggering 80 per cent of these platform disccussions lead to specific government action.

It is only by making sure the issues are broadly understood will we be able to ensure that differing perspectives are heard, including the insights of those who are at the sharp end, such as those rendered jobless by automation.



AI is already working with artists and musicians to create new content
GETTY IMAGES

5. Search for a truce in the AI arms race

There is a powerful global dimension to the race for AI superiority, with China vying with the US for leadership. Any state that decelerates its development of AI in the name of ethics risks handing a devastating advantage to its less scrupulous rivals.

ADVERTISEMENT

So we face the daunting challenge of creating a global system for the regulation of AI, akin to the global architecture for nuclear arms control, at a time of confrontation between the great powers.

The prospects of a global consensus may seem bleak, but we have to try. We urgently need to develop minimally adequate standards for AI regulation that reflect what different countries and ideologies will be able to accept.

This means giving up on the best to avoid the very worst. A first step might be to reach a consensus among leading democratic states on guidelines for developing and deploying AI and then to see how far we need to dilute that consensus to bring China and other key states on board.

It's not impossible that Britain, intermediate geographically and ideologically between a libertarian US and a Napoleonic EU, may have a constructive role to play in this urgent task.

6. Help it become a force for good

Finally, let us not lose sight of the extraordinary potential for AI to enrich our lives, our societies and economies. AI will continue to revolutionise medicine, from the discovery of new drugs to early diagnosis of disease, from understanding our biology to managing every sort of medical surveillance. AI will help us understand climate change and work with humans to find ways to ameliorate its impact. It is already working with artists and musicians to create new content.

AI can empower us rather than diminish us if we are all part of a conversation about the values we wish to imbue in our systems and the regulatory environment in which they exist.

John Tasioulas is a professor of ethics and legal philosophy at the University of Oxford and Director of the Institute for Ethics in AI

ADVERTISEMENT

Nigel Shadbolt is a professor of computer science at the University of Oxford and distinguished senior scientist at the Institute for Ethics in AI

Technology
Artificial Intelligence