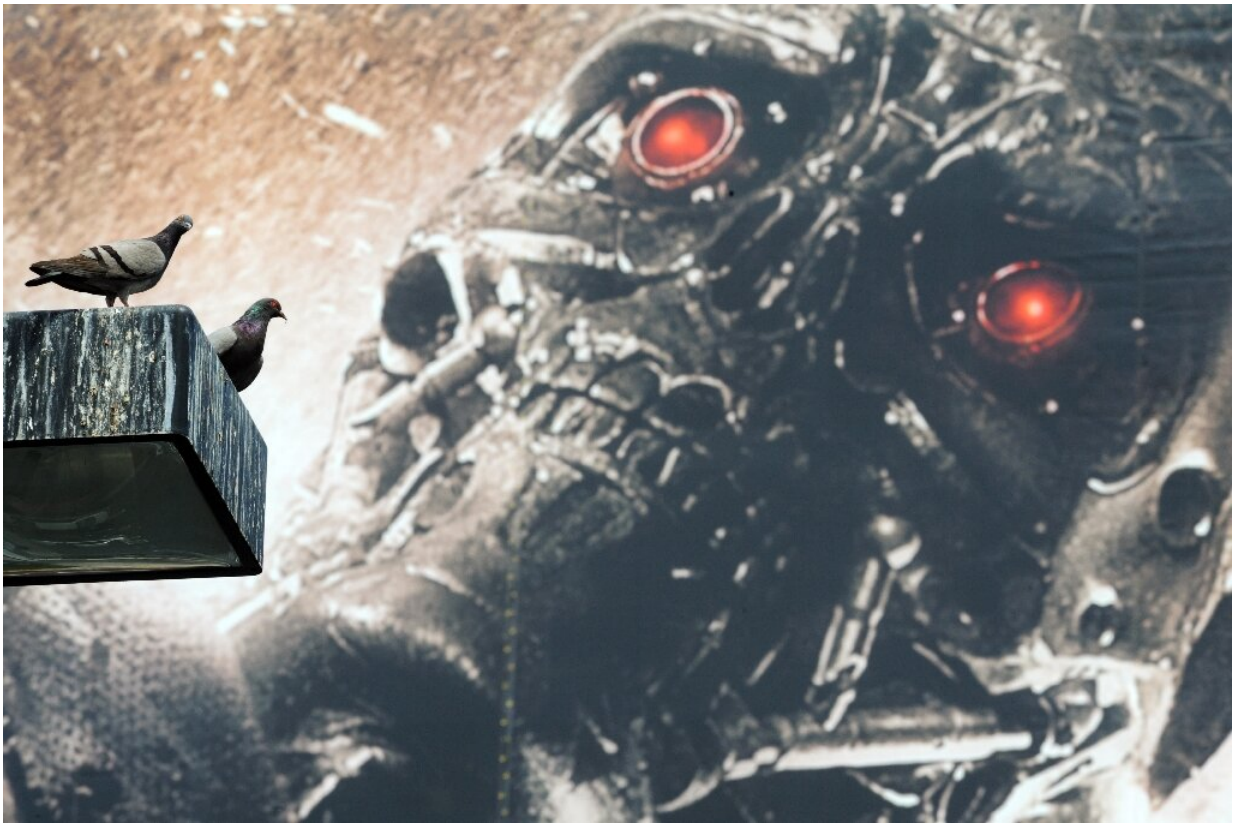
Will AI really destroy humanity?

June 27 2023, by Joseph BOYLE



The Stop Killer Robots group has explicitly dismissed the Terminator scenario.

The warnings are coming from all angles: artificial intelligence poses an existential risk to humanity and must be shackled before it is too late.

But what are these disaster scenarios and how are machines supposed to wipe out humanity?

Paperclips of doom

Most disaster scenarios start in the same place: machines will outstrip human capacities, escape human control and refuse to be switched off.

"Once we have machines that have a self-preservation goal, we are in trouble," AI academic Yoshua Bengio told an event this month.

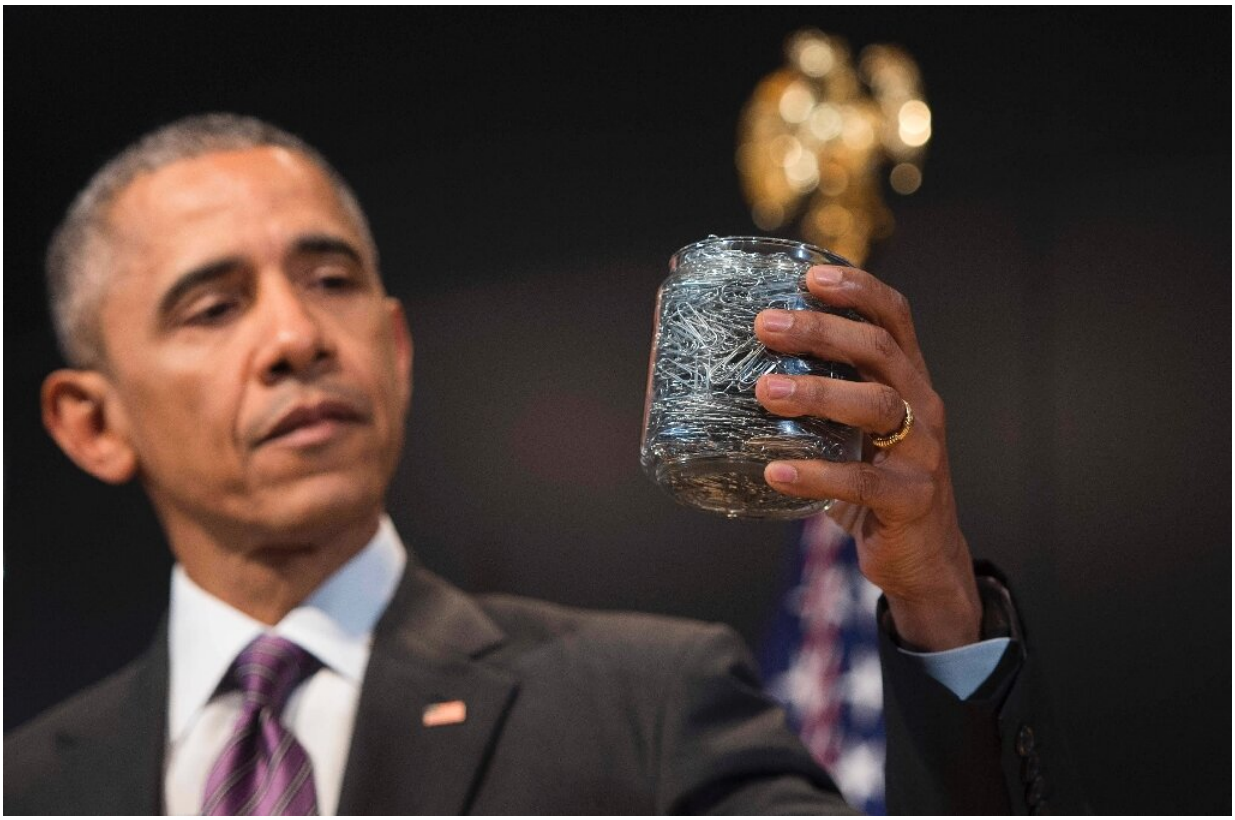
But because these machines do not yet exist, imagining how they could doom humanity is often left to philosophy and science fiction.

Philosopher Nick Bostrom has written about an "intelligence explosion" he says will happen when superintelligent machines begin designing machines of their own.

He illustrated the idea with the story of a superintelligent AI at a paperclip factory.

The AI is given the ultimate goal of maximizing paperclip output and so "proceeds by converting first the Earth and then increasingly large chunks of the observable universe into paperclips".

Bostrom's ideas have been dismissed by many as science fiction, not least because he has separately argued that humanity is a computer simulation and supported theories close to eugenics.



Philosopher Nick Bostrom dreamt up the idea of a superintelligent AI converting the Earth into paperclips.

He also recently apologized after a racist message he sent in the 1990s was unearthed.

Yet his thoughts on AI have been hugely influential, inspiring both Elon Musk and Professor Stephen Hawking.

The Terminator

If superintelligent machines are to destroy humanity, they surely need a physical form.

Arnold Schwarzenegger's red-eyed cyborg, sent from the future to end human resistance by an AI in the movie "The Terminator", has proved a seductive image, particularly for the media.

But experts have rubbished the idea.

"This science fiction concept is unlikely to become a reality in the coming decades if ever at all," the Stop Killer Robots campaign group wrote in a 2021 report.

However, the group has warned that giving machines the power to make decisions on life and death is an existential risk.

Robot expert Kerstin Dautenhahn, from Waterloo University in Canada, played down those fears.

AI scholar Stuart Russell reckons real killer robots will be small, airborne and come in swarms.

She told AFP that AI was unlikely to give machines higher reasoning capabilities or imbue them with a desire to kill all humans.

"Robots are not evil," she said, although she conceded programmers could make them do evil things.

Deadlier chemicals

A less overtly sci-fi scenario sees "bad actors" using AI to create toxins or new viruses and unleashing them on the world.

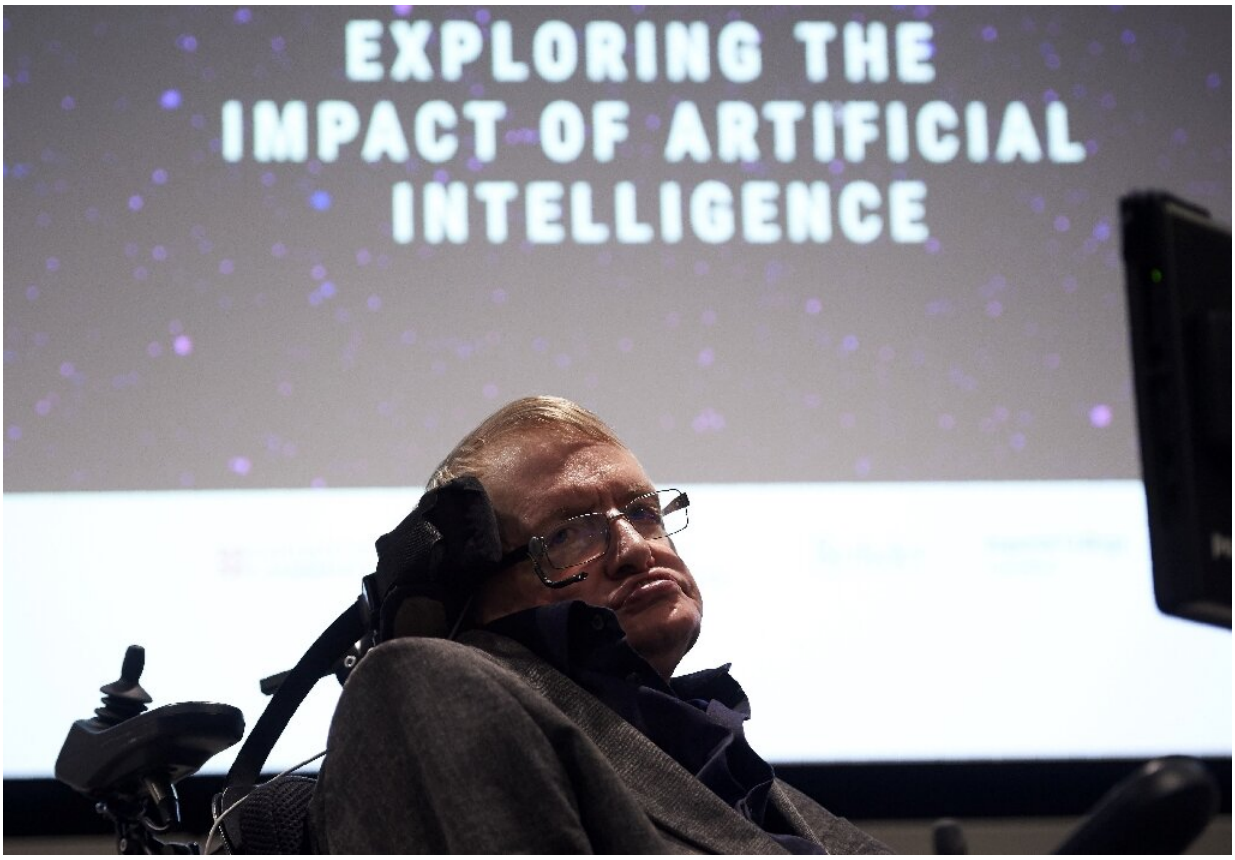
Large language models like GPT-3, which was used to create ChatGPT, it turns out are extremely good at inventing horrific new chemical agents.

A group of scientists who were using AI to help discover new drugs ran an experiment where they tweaked their AI to search for harmful molecules instead.

They managed to generate 40,000 potentially poisonous agents in less than six hours, as reported in the Nature Machine Intelligence journal.

AI expert Joanna Bryson from the Hertie School in Berlin said she could imagine someone working out a way of spreading a poison like anthrax more quickly.

"But it's not an existential threat," she told AFP. "It's just a horrible, awful weapon."



Stephen Hawking argued in 2014 that at some point in the future superintelligent machines will surpass human abilities and ultimately our species will no longer be able to compete.

Species overtaken

The rules of Hollywood dictate that epochal disasters must be sudden, immense and dramatic—but what if humanity's end was slow, quiet and not definitive?

"At the bleakest end our species might come to an end with no successor," philosopher Huw Price says in a promotional video for Cambridge University's Centre for the Study of Existential Risk.

But he said there were "less bleak possibilities" where humans augmented by advanced technology could survive.

"The purely biological species eventually comes to an end, in that there are no humans around who don't have access to this enabling technology," he said.

The imagined apocalypse is often framed in evolutionary terms.

Stephen Hawking argued in 2014 that ultimately our species will no longer be able to compete with AI machines, telling the BBC it could "spell the end of the human race".

Geoffrey Hinton, who spent his career building machines that resemble the human brain, latterly for Google, talks in similar terms of "superintelligences" simply overtaking humans.

He told US broadcaster PBS recently that it was possible "humanity is just a passing phase in the evolution of intelligence".

Citation: Will AI really destroy humanity? (2023, June 27) retrieved 29 April 2024 from https://techxplore.com/news/2023-06-ai-destroy-humanity.html