

Is Stephen Hawking right? Could AI lead to the end of humankind?

December 4 2014, by David Dowe



British astrophysicist Professor Stephen Hawking with his new Intel-created communications platform. Credit: EPA/Andy Rain

The famous theoretical physicist, [Stephen Hawking](#), has revived the debate on whether our search for improved artificial intelligence will one day lead to thinking machines that will take over from us.

The British scientist made the claim during a wide-ranging interview with the BBC. Hawking has the [motor neurone disease, amyotrophic lateral sclerosis](#) (ALS), and the interview touched on new technology he is using to help him communicate.

It works by modelling his previous word usage to predict what words he will use next, similar to predictive texting available on many smart phone devices.

But Professor Hawking also mentioned his concern over the development of [machines](#) that might surpass us.

"Once humans develop artificial intelligence, it would take off on its own and re-design itself at an ever increasing rate," he [reportedly told](#) the [BBC](#).

"The development of full artificial intelligence could spell the end of the human race."

Could thinking machines take over?

I appreciate the issue of computers taking over (and one day ending humankind) being raised by someone as high profile, able and credible as Prof Hawking – and it deserves a quick response.

The issue of [machine intelligence](#) goes back at least as far as the British code-breaker and father of computer science, [Alan Turing](#) in [1950](#), when he considered the question: "Can machines think?"

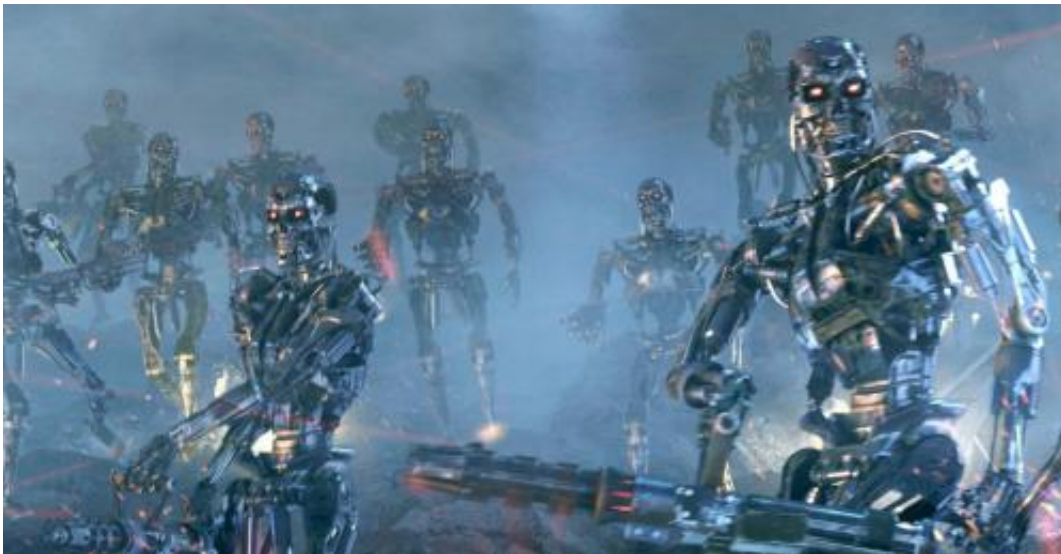
The issue of these [intelligent machines](#) taking over has been discussed in one way or another in a variety of popular media and culture. Think of the movies [Colossus – the Forbin project](#) (1970) and [Westworld](#) (1973), and – more recently – Skynet in the 1984 movie [Terminator](#) and sequels, to name just a few.

Common to all of these is the issue of delegating responsibility to machines. The notion of the [technological singularity](#) (or machine super-intelligence) is something which goes back at least as far as [artificial](#)

[intelligence pioneer](#), [Ray Solomonoff](#) – who, in [1967](#), warned:

Although there is no prospect of very intelligent machines in the near future, the dangers posed are very serious and the problems very difficult. It would be well if a large number of intelligent humans devote a lot of thought to these problems before they arise.

*It is my feeling that the realization of artificial intelligence will be a sudden occurrence. At a certain point in the development of the research we will have had no practical experience with machine intelligence of any serious level: a month or so later, we will have a **very** intelligent machine and all the problems and dangers associated with our inexperience.*



When Skynet took over in the Terminator movies it sent forth killing machines to wipe out humans. Credit: EPA PHOTO/EFE/Columbia TriStar/Robert Zucker

As well as giving this variant of Hawking's warning back in 1967, in [1985 Solomonoff](#) endeavoured to give a time scale for the technological singularity and reflect on social effects.

I share the concerns of [Solomonoff](#), [Hawking](#) and others regarding the consequences of faster and more intelligent machines – but American author, computer scientist and inventor, [Ray Kurzweil](#), is one of many seeing the benefits.

Whoever might turn out to be right (provided our planet isn't destroyed by some other danger in the meantime), I think [Solomonoff](#) was prescient in [1967](#) in advocating we devote a lot of thought to this.

Machines already taking over

In the meantime, we see increasing amounts of responsibility being delegated to machines. On the one hand, this might be hand-held calculators, routine mathematical calculations or global positioning systems (GPSs).

On the other hand, this might be systems for [air traffic control](#), guided missiles, driverless trucks on mine sites or the recent trial appearances of driverless cars on our roads.

Humans delegate responsibility to machines for reasons including improving time, cost and accuracy. But nightmares that might occur regarding damage by, say a driverless vehicle, would include legal, insurance and attribution of responsibility.

It is argued that computers might take over when their [intelligence](#) supersedes that of humans. But there are also other risks with this delegation of responsibility.

Mistakes in the machines

Some would contend that the stock market crash of 1987 was largely due

to [computer trading](#).

There have also been power grid closures due to computer error. And, at a lower level, my intrusive spell checker sometimes "corrects" what I've written into something potentially offensive. Computer error?

Hardware or software glitches can be hard to detect but they can still wreak havoc in large-scale systems – even without hackers or malevolent intent, and probably more so with them. So, just how much can we really trust machines with large responsibilities to do a better job than us?

Even without computers consciously taking control, I can envisage a variety of paths whereby computer systems go out of control. These systems might be so fast with such small componentry that it might be hard to remedy and even hard to turn off.

Partly in the spirit of [Solomonoff's 1967](#) paper, I'd like to see scriptwriters and [artificial intelligence](#) researchers collaborating to set out such scenarios – further stimulating public discussion.

As but one possible scenario, maybe some speech gets converted badly to text, worsened in a bad automatic translation, leading to a subtle corruption of machine instructions, leading to whatever morass.

A perhaps related can of worms might come from faster statistical and machine learning analysis of big data on human brains. (And, as some would dare to add, are we humans the bastions of all that is good, moral and right?)

As Solomonoff said in [1967](#), we need this public discussion – and, given the stakes, I think we now need it soon.

This story is published courtesy of [The Conversation](#) (under Creative

Commons-Attribution/No derivatives).

Source: The Conversation

Citation: Is Stephen Hawking right? Could AI lead to the end of humankind? (2014, December 4) retrieved 25 April 2024 from

<https://phys.org/news/2014-12-stephen-hawking-ai-humankind.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.