The rise of artificial intelligence – and its threat to humanity

General-purpose artificial intelligence could eventually mark the end of human innovation, writes Cliff Saran

Autonomous cars, automated trading and smart cities are among the great promises of machine intelligence. But artificial intelligence (AI) promises much more – including being man's best friend.

BigDog was a robot developed in 2008, funded by the US Defense Advanced Research Projects Agency and the US Army Research Laboratory's RCTA programme. It was designed to walk and climb – skills humans master instinctively at an early age, but which cannot easily be programmed into a machine. Instead, researchers applied AI techniques to enable it to "learn".

Imagine a computer that can think better than humans; that can make profound cognitive decisions at lightning speed. Such a machine could better serve mankind. But would it?

"AI that can run 1,000 times faster than humans can earn 1,000 times more than people," according to Stuart Armstrong, research fellow at the Future of Humanity Institute. "It can make 100 copies of itself." This ability to think fast and make copies of itself is a potent combination – and one than could have a profound effect on humanity.

"With human-level intelligence, plus the ability to copy, it could hack the whole internet," Armstrong warned. And if this general-purpose AI had a body, he said, "it could walk into a bar and walk out with all the girls or guys".

> **"Were it programmed to prevent all human suffering, the solution could be to kill all humans"**
>
> Stuart Armstrong,
> Future of Humanity Institute

But far from being a super hacker – or master pick-up artist – Armstrong argues that, were such machines to become powerful, the world would resemble their preferences. For instance, he said they could boost their own algorithms.

Beware of extreme machine intelligence

Socially aware, general-purpose AI could scan the web for information and, by reading human facial expressions, it could deliver targeted speeches better than any political leader, he said.

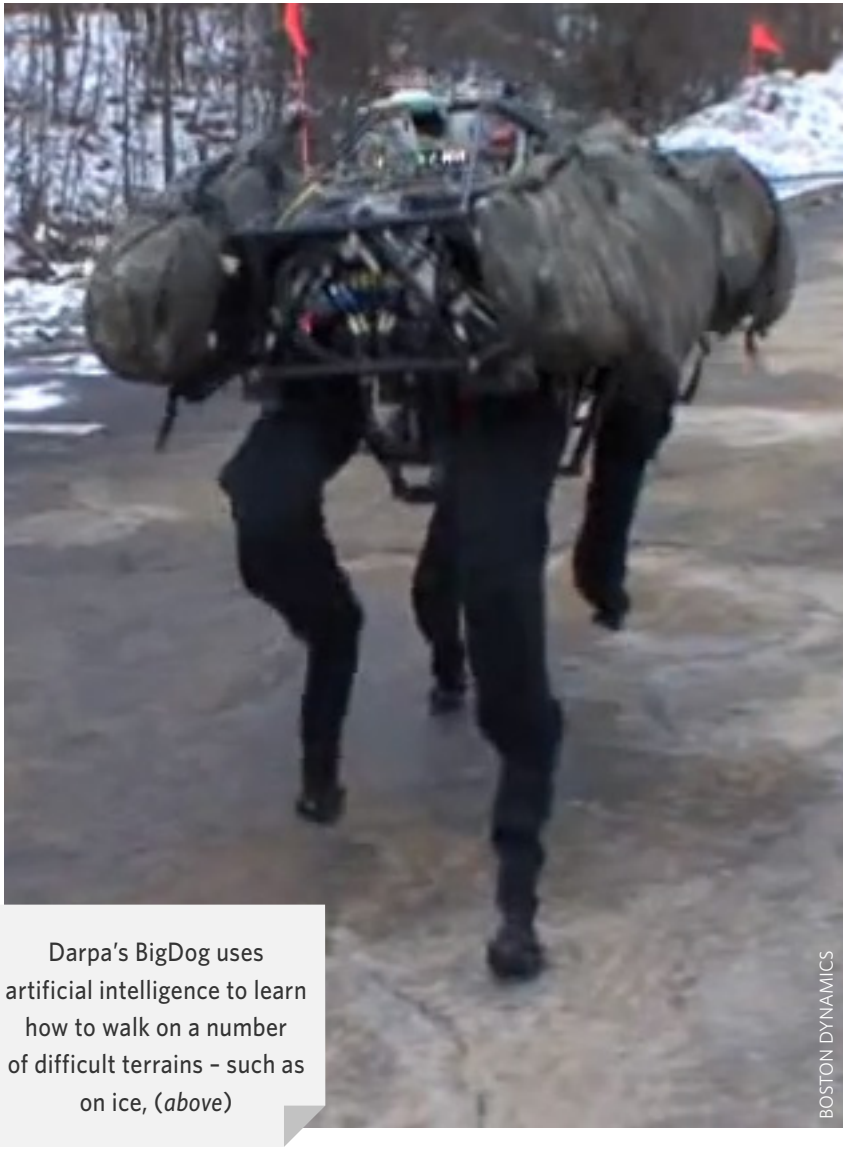
Taken to the extreme, Armstrong warned that it is difficult to specify a goal that is safe: "If it were programmed to prevent all human suffering, the solution could be to kill all humans."

In his book *Smarter than us*, Armstrong laid down a few points humans should consider about general-purpose AI: "Never trust an entirely super-intelligent AI. If it doesn't have your best interests at heart, it'll find a way to obey all its promises while still destroying you."

Such AI remains a long way off. Armstrong's closest estimate of when such intelligence could be developed fell somewhere between five and 150 years' time. But it is a hot topic, and London-based DeepMind recently demonstrated how a machine used reinforcement, learning to take what it had learned from playing a singe Atari 2600 game and applying it to other computer games.

DeepMind is not general AI, according to Armstrong. It is narrow AI – a form of artificial intelligence that can undertake tasks people once said would not be possible without general-purpose AI. IBM's Watson – which won US TV game show *Jeopardy* – and Google's driverless car are applications of narrow AI.

Gartner analyst Steve Prentice said narrow AI is a machine that does one task particularly well: "The variables have to be limited, and it follows a set of rules." For instance, he said an autonomous vehicle could be programmed in a way that could prevent cycle road deaths.



BOSTON DYNAMICS

Darpa's BigDog uses artificial intelligence to learn how to walk on a number of difficult terrains – such as on ice, (*above*)

Robots could rule the world

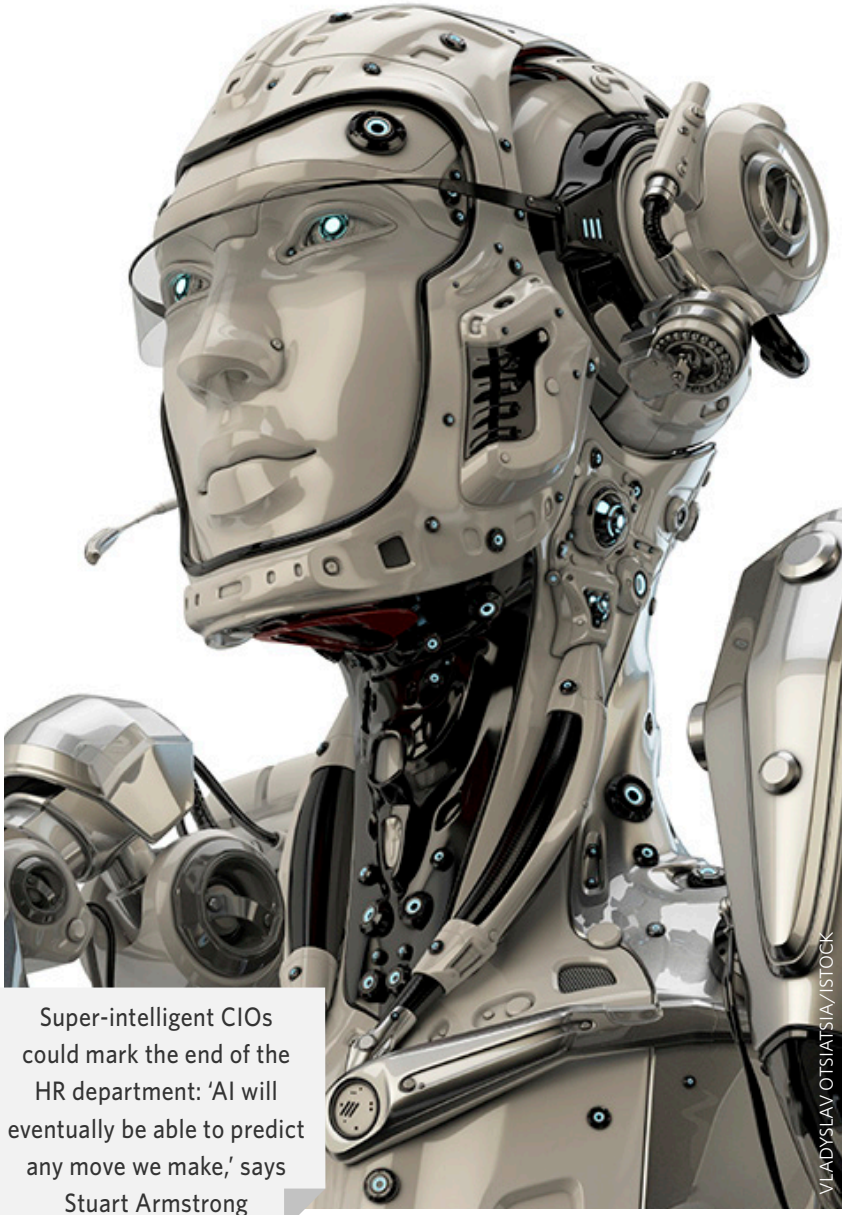
In the Gartner report, *When smart things rule the world*, Prentice argued the case for CIOs to start thinking about the business impact of smart machines that exhibit AI behaviour. He noted: "Advanced capabilities afforded by artificial intelligence will enhance today's smart devices to display goal-seeking and self-learning behaviour, rather than a simple sense and respond."

For CIOs, Prentice regards autonomous business as a logical extension of current automated processes to increase efficiency, rather than simply to replace a human workforce. "For most people, AI is slanted to what you see on screen. But from a business perspective, we are far away from this in reality," he said.

In fact, he said there is no reason why a super-intelligent AI machine could not act like a CEO or manager, directing humans to do tasks where creativity or manual dexterity is important.

This may sound like a plot from Channel 4 sci-fi drama *Humans* but, as Armstrong observed in *Smarter than us*: "Even if the AI is nominally under human control, even if we can reprogram it or order it around, such theoretical powers will be useless in practice. This is because the AI will eventually be able to predict any move we make and could spend a lot of effort manipulating those who have 'control' over it."

So back to man's best friend. Armstrong is not afraid of the metal-clad robot with an Austrian accent that Arnold Schwarzenegger depicted in *The Terminator*. For him, a super-intelligent machine taking the form of a dog and biting the proverbial hand that feeds it is a far more plausible way in which machines could eventually rule the world. ■



VLADYSLAV OTSIATSIA/ISTOCK

Super-intelligent CIOs could mark the end of the HR department: 'AI will eventually be able to predict any move we make,' says Stuart Armstrong

› *Online grocer Ocado is supporting a civilian robot research initiative*