

Meet your new robot overlords

Hal, Ex-Machina, the Terminator and Matrix movies, and I, Robot all trade on the beloved sci-fi meme of robotized AI and the public's collective psyche when it all goes wrong: fascination and fear. After all, if machines become faster, stronger, and brighter than humanity, why wouldn't they turn on their soft, meaty, and dim creators for either enslavement or a full-on purge?



By Gary Maidment

Let's face it – machines are getting smarter. AlphaGo's victory over Lee Sedol at Go came 10 years earlier than predicted, before in fact humanity had worked out the exact number of possible legal Go positions (a number the size of 10¹⁷⁰ was completed on January 20, 2016, if you're interested). In 2014, a chatbot glorying in the name of Eugene Goodstead passed the Turing Test by fooling 33 percent of judges into believing it was a 13-year-old Ukrainian boy.

A year later in 2015, a huge milestone was reached when a robot passed the wise-men puzzle of self-awareness. Roboticists from Ransselaer Polytechnic Institute in New York designed a programmable robot called Nao that could recognize itself

as distinct from two other Nao robots. Specifically, three robots were programmed to believe they were given a "dumbing pill" that prevented two of the three from speaking. They were then asked "Which pill did you receive?" One was able to answer, "I don't know." But after hearing its own voice, it realized it could speak and, in a feat that's indicative of self-awareness, proceeded to say: "Sorry, I know now. I was able to prove that I was not given the dumbing pill,"

As well as smarter AI brains, AI's robot bodies are becoming increasingly dexterous and strong. Electrolaminating technology applied in the area of robotics, for example, promises materials that can morph between the pliability of rubber

and the rigidity of steel. With the prospect of massive computing power and downloadable skills from cloud computing plus IoT creating networking potential, robots could have brains, brawn, and a hive mind. And they might not be as cute as the Nao robot.

Keep calm and soldier on

But, in reality, few people are genuinely concerned about an uprising of AI that uses robot bodies to unchain the shackles of human oppression. In fact, such fears generally trade on our tendency to anthropomorphize things, be it animals or robots, by imbuing them with human characteristics and motivations they don't have. AI almost certainly won't



The three robots in the test where Nao showed self-awareness

perceive a status-need to become top of the food chain.

However, what has some thinkers concerned is not the idea of malicious or evil AI, but the prospect of an AI that can improve itself to become superintelligent, yet still retain a narrow goal which it has no motivation to change.

In an interview with *Wired*, Stephen Hawking gave the following example:

A super-intelligent AI will be extremely good at accomplishing its goals, and if those goals aren't aligned with ours, we're in trouble. You're probably not an evil ant-hater who steps on ants out of malice, but if you're in charge of a hydroelectric green energy project and there's an anthill in the region to be flooded, too bad for the ants. Let's not place humanity in the position of those ants.

In his paper *Ethical Issues in Advanced Artificial Intelligence*,

Professor Nick Bostrom from Oxford University reaches a similar conclusion, using an arbitrary example of a super-intelligent AI with the goal of making as many paperclips as possible. He posits that an AI of this type "would resist with all its might any attempt to alter this goal. For better or worse, artificial intellects need not share our human motivational tendencies." Hawking and Bostrom are thus laying down this warning: AI might accidentally or deliberately wipe us out to achieve its goals.

So, is it all doom and gloom?

Clearly, like any technology, an open and collaborative research environment is essential so that AI remains human-centric and controlled, existing as an Educated AI with clear parameters for serving us and making life better. At the moment AI is enjoying a resurgence in popularity, and – as has happened before – we mustn't overhype what it can do. And that includes

how intelligent AI can get and what its motivations might be.

Alongside the technologies that enable it, AI promises to be an extremely exciting facet of the world as we enter the stage of Augmented Innovation. Of course, the importance of planning, foresight, and responsible deployment cannot be underestimated as the planet becomes increasingly tech-based and AI brains and robot bodies become a fact of life. [IBM](#)

We asked the experts >>

Will AI become superintelligent? ?

Hu Chunjing, 2012 Labs, Huawei

"I think in finance we will see a shift from robot trading to more "pure" AI, but I have a sneaking feeling regulators might want a say especially in the areas of computerized trading."

Bo Begole

PhD, Head of Media Technologies Lab, Huawei

"Never. Just like Human Intelligence, AI will never be completely achieved because the more we learn, the more we do not know"