

2 December 2014 Last updated at 12:03 ET

Does rampant AI threaten humanity?

By Mark Ward

Technology correspondent, BBC News

Pity the poor meat bags. They are doomed if a growing number of scientists, engineers and artists are to be believed.

Prof Stephen Hawking has joined a roster of experts worried about what follows when humans build a device or write some software that can properly be called intelligent.

Such an artificial intelligence (AI), he fears, could spell the end of humanity.

Similar worries were voiced by Tesla boss Elon Musk in October. He declared rampant AI to be the "biggest existential threat" facing mankind. He wonders if we will find our end beneath the heels of a cruel and calculating artificial intelligence.

So too does the University of Oxford's Prof Nick Bostrom, who has said an AI-led apocalypse could engulf us within a century.

Google's director of engineering, Ray Kurzweil, is also worried about AI, albeit for more subtle reasons. He is concerned that it may be hard to write an algorithmic moral code strong enough to constrain and contain super-smart software.

Many films, such as The Terminator movies, 2001, The Matrix, Blade Runner, to mention a few, pit puny humans against AI-driven enemies.

More recently, Spike Jonze's Her involved a romance between man and operating system, Alex Garland's forthcoming Ex Machina debates the humanity of an android and the new Avengers movie sees superheroes battle Ultron - a super-intelligent AI intent on extinguishing mankind. Which it would do with ease were it not for Thor, Iron Man and their super-friends.

Even today we are getting hints about how paltry human wits can be when set against computers who throw all their computational horsepower at a problem. Chess computers now routinely beat all but the best human players. Complicated mathematics is a snap to as lowly a device as the smartphone in your pocket.

IBM's Watson supercomputer took on and beat the best players of US TV game show Jeopardy. And there are many, many examples of computers finding novel and creative solutions to problems across diverse fields that, before now, never occurred to us humans.

The machines are slowly but surely getting smarter and the pursuits in which humans remain champions are diminishing.

Death by drone

But is the risk real? Once humans code the first genuinely smart computer program that then goes on to develop its smarter successors, is the writing on the wall for humans?

Maybe, said Neil Jacobstein, AI and robotics co-chairman at California's Singularity University.

"I don't think that ethical outcomes from AI come for free," he said, adding that work now will significantly improve our chances of surviving the rise of rampant AI.

What we must do, he said, is consider the consequences of what we were creating and prepare our societies and institutions for the sweeping changes that might arise.

"It's best to do that before the technologies are fully developed and AI and robotics are certainly not fully developed yet," he said. "The possibility of something going wrong increases when you don't think about what those potential wrong things are."

"I think there is a great opportunity for us to be proactive about anticipating those possible negative risks and doing our best to develop redundant, layered thoughtful controls for those risks," he told the BBC.

So far, said Murray Shanahan, professor of cognitive robotics at Imperial College, those actively working on AI were not really putting in place safety systems to stop their creations running amok.

"The AI community does not think its a substantial worry," he said, "whereas the public does think it's much more of an issue."

"The right place to be is probably in-between those two extremes," he said, adding: "There's no need for panic right now.

"I do not think we are about to develop human-level AI within the next 10-20 years," he said. "On the other hand its probably a good idea for AI researchers to start thinking about the issues that Stephen Hawking and others have raised."

And, said Prof Shanahan, the greatest obstacle to developing those genuinely smart machines had yet to be overcome - how we actually create machine-based intelligence.

"We do not really know yet whether the best way is to copy nature or start from scratch," said Prof Shanahan.

For science-fiction author Charles Stross, the dangers inherent in artificially smart systems do not arise because they will out-think us or suddenly realise they can please themselves rather than their human masters.

"Nobody wants an AI that will set its own goals because the probable outcome is that it will decide to do the AI equivalent of sacking out on the sofa with a bowl of chips and the cable TV controller rather than doing whatever it is that we consider to be useful," he said.

A glance at all the work being done on AI right now shows that much of it is concentrating on systems that specifically lack the autonomy and consciousness that could spell problems for us humans.

The AI's we were getting now and which were likely to appear in the future might be dangerous, Stross said, but only because of the people they served.

"Our biggest threat from AI, as I see it, comes from the consciousnesses that set their goals," he said.

"Drones don't kill people - people who instruct drones to fly to grid coordinates (X, Y) and unleash a Hellfire missile kill people," he said. "It's the people who control them whose intentions must be questioned.

"We're already living in the early days of the post-AI world, and we haven't recognised that all AI is is a proxy for our own selves - tools for thinking faster and more efficiently, but not necessarily more benevolently," he said.

More videos from the BBC