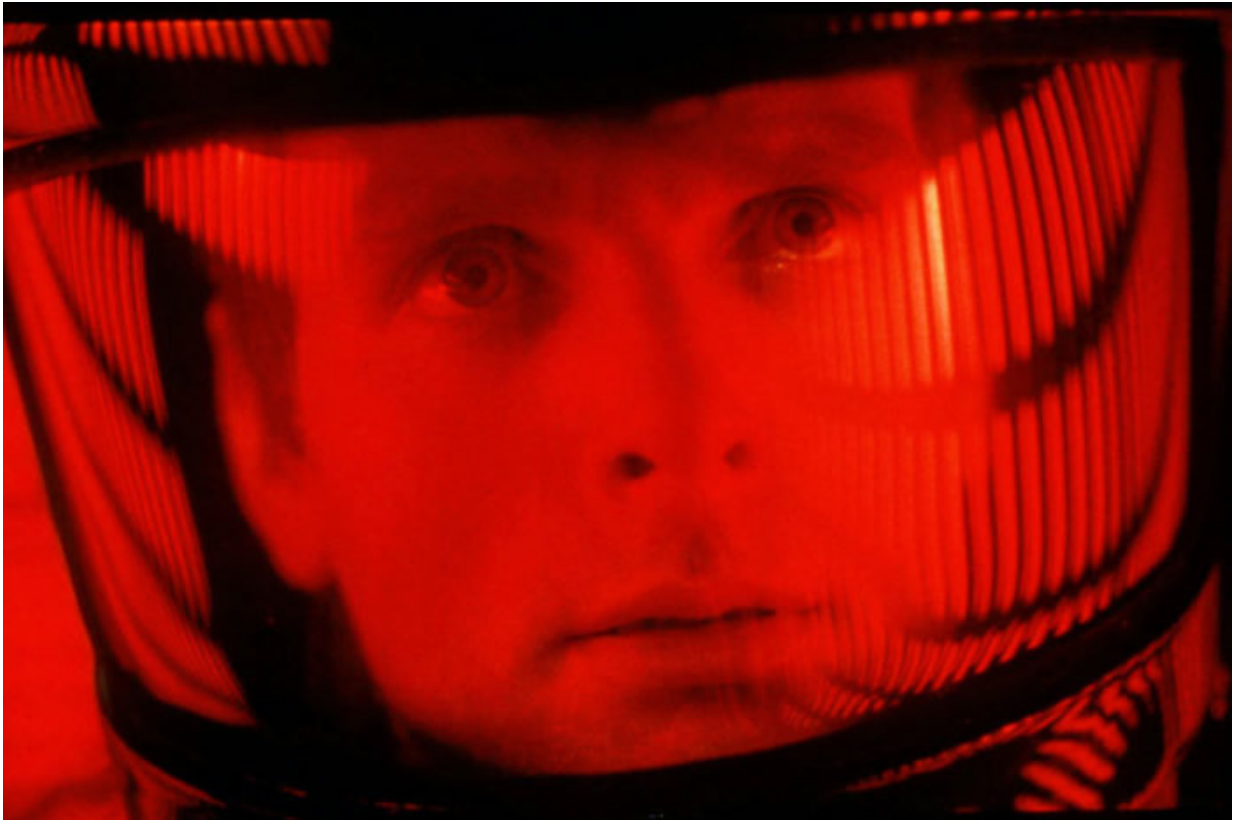


Kubrick's AI nightmare, 50 years later

April 13 2018



Credit: University of Western Ontario

As David Bowman – the surviving crew member aboard the Discovery One spacecraft in Stanley Kubrick's 2001: A Space Odyssey – disassembles HAL 9000, the sentient computer pleads in an affectless, monotone voice:

"I'm afraid, Dave."

"Dave, my mind is going. I can feel it."

As HAL's consciousness – or rather, his logic – fades, he dies singing Daisy Bell, the first song 'sung' by a real-world computer. With the threat removed, all is seemingly right again.

Celebrating its 50th anniversary this month, Kubrick's masterpiece has cast a shadow over the genre since its premiere. Its influence extends beyond depictions of space and space travel, touching more than Star Wars, Alien or Blade Runner.

For example, its effect on our vision of artificial intelligence (AI) is palpable.

Think of Amazon's Alexa, who, like HAL, listens to whatever you say.

But now, five decades later, have we evolved past Kubrick's nightmare of a sentient, threatening machine? How has our understanding of, and relationship to, AI changed? Do we have a reason to fear the machines we program?

For Catherine Stinson, who recently completed a postdoctoral fellowship at Western's Rotman Institute of Philosophy, Kubrick's vision, while much different from the present state of AI, is still a looming threat. The threat, however, is not the machine.

"People thought about AI a lot differently back then, the danger being it was going to be an agent who would act differently than us, with different priorities than what we have," she said.

"That is less the worry now. It's not going to be the one-on-one

interactions (with a sentient machine) that we don't know how to deal with. It's going to be something we've put all our evil into, and now it's off doing things that are an extension of the problems of humans – but on a grander scale we couldn't have imagined. It's not so much machines are different from us – it's they are reproducing the problems of humans."

Part of the issue, Stinson explained, is humans are the ones programming AI. How can humans program ethical machines? Can machines be ethical? We see ourselves as being competent in making ethical decisions because we decide between right and wrong on a regular basis, she said. We rely on an instinct we know right from wrong in day-to-day situations.

"But in more complicated situations that come up – like self-driving cars – it's really difficult, even for someone who does have training in ethics, to design what the right thing to build into it is," Stinson noted.

For instance, should the car avoid crashing into a pedestrian, even if it is going to lead to the death of the driver? How do you weigh the two different lives at risk? Do you program the car to save the occupants of the vehicle or those with whom it might collide?

"I don't know how to make that kind of decision. I don't know that that decision is something the average person knows how to make. It needs a more careful approach and someone with more expertise needs to be involved. But it's hard to see that there is that need, because everyone thinks they are an expert," Stinson added.

Individuals taking engineering and technology courses should be trained in ethics, she added. Barring that, companies working in AI could benefit from an in-house ethicist. Academic institutions are increasingly requiring engineers and computer scientists to take courses that touch on

the subject. Although the question of 'ethical machines' is up for debate, the simple fact we can program them to perform acts that are right or wrong involves them in an "ethical game," Stinson said.

"Maybe we could program a machine to do the right thing more often than we would. But is there reason to fear? Sure. There are machines being used in the justice system in the United States, making decisions that maybe aren't the right ones. We're not sure how they are making those decisions and there's no accountability to whose fault it is if they make the wrong decision," she noted.

For sentencing in particular, there are AI programs that help judges decide on what the right sentence should be for someone convicted of a crime. The algorithm is designed to make sentencing less biased by taking into account factors from the person's past, what kind of neighbourhood they grew up in, what kind of people they knew, prior arrests, age of first involvement with police, etc.

All of those things are not neutral pieces of information, Stinson said. Such AI programs have been criticized for reinforcing the stereotypes they were designed to avoid.

"We don't know what the dangers are. Part of worrying about the dangers is trying to predict what those might be, and to decide on what we value, and what kind of things we want to have happen, for the sake of convenience," Stinson said.

Tim Blackmore, a professor in the Faculty of Information and Media Studies, has taught *2001: A Space Odyssey* to students for more than a decade. He echoed Stinson, noting the dangers of AI lie in the human element at play. For him, whatever form it takes in films or books, AI has always been an extension of the human.

"Thinking machines are often portrayed as cognisant of their own existence and aware of existential issues. They are one of the many mirrors humans use to reflect what it is to be human," Blackmore said.

And that's the nightmare.

"Until now, it's been a 'machine that rules the world' kind of nightmare. That comes out of the 1960s and is shaped very much by Vietnam, as well as the idea these mainframes, these big machines, were part of a worldview that was running us into an inhuman, determinist way of living that would lead to genocides," he explained.

But the threat today lies not in our vision of AI as some machine from the future that can outperform or conquer us.

"We much less imagine WALL-E – the helper machine. But that's much more it. It's not the machines that are a problem; it's the humans. People do bad things," Blackmore noted, adding he is nervous about the "helper" [machines](#) we blindly embrace.

"I'm worried about these disks and cylinders or whatever Amazon, Google or Facebook want to jam into our home next. People want this; it's a gadget and it's cool because it's so hard to pick up your mobile phone and type something into it or speak into it. We're going into the trough and we suck that stuff up, and then we're going to have terabytes of data flying into pools where they could be scrubbed for everything. That data can be manipulated by AI agents who will be better and better at looking for how to game [human](#) beings," he continued.

"How this technology will develop so people can push people around – that is what tends to be bad news for us. The robot uprising is lower on my list."

Provided by University of Western Ontario

Citation: Kubrick's AI nightmare, 50 years later (2018, April 13) retrieved 5 July 2024 from <https://phys.org/news/2018-04-kubrick-ai-nightmare-years.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.