

He May Be Your Best Hope Against Killer Robots

New York Times 7/30/19

Autonomous weapons are closer than we think, this leading A.I. expert believes.

By CLAUDIA DREIFUS

Toby Walsh, a professor at the University of New South Wales in Sydney, is one of Australia's leading experts on artificial intelligence. He and other experts have released a report outlining the promises, and ethical pitfalls, of the country's embrace of A.I.

Recently, Dr. Walsh, 55, has been working with the Campaign to Stop Killer Robots, a coalition of scientists and human rights leaders seeking to halt the development of autonomous robotic weapons.

We spoke briefly at the annual meeting of the American Association for the Advancement of Science, then for two hours via telephone. Below is an edited version of those conversations.

You are a scientist and an inventor. How did you become an activist in the fight against 'killer robots'?

It happened incrementally, beginning around 2013. I had been doing a lot of reading about robotic weaponry. I realized how few of my artificial intelligence colleagues were thinking about the dangers of this new class of weapons. If people thought about them at all, they dismissed killer robots as something far in the future.

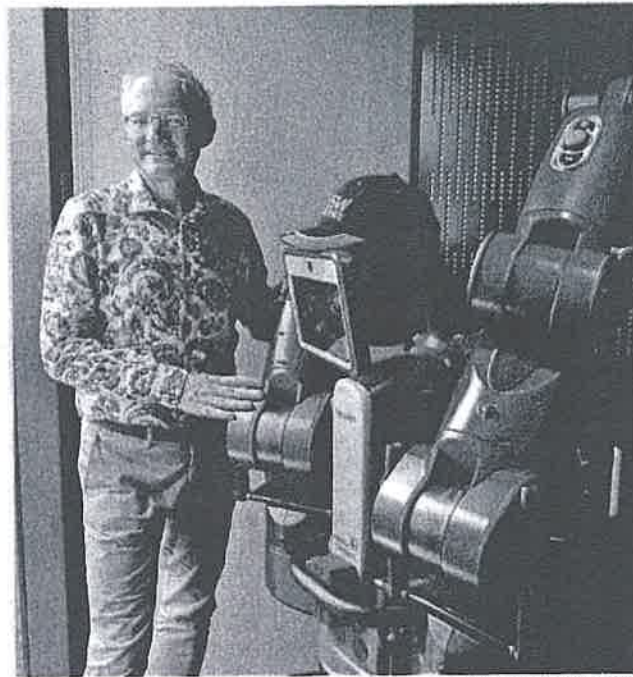
From what I could see, the future was already here. Drone bombers were flying over the skies of Afghanistan. Though humans on the ground controlled the drones, it's a small technical step to render them autonomous.

So in 2015, at a scientific conference, I organized a debate on this new class of weaponry. Not long afterward, Max Tegmark, who runs M.I.T.'s Future of Life Institute, asked if I'd help him circulate a letter calling for the international community to pass a pre-emptive ban on all autonomous robotic weapons.

I signed, and at the next big A.I. conference, I circulated it. By the end, we had over 5,000 signatures — including Elon Musk, Daniel Dennett, Steve Wozniak.

What was your argument?

That you can't have machines deciding



CLAUDE SEWELL FOR THE NEW YORK TIMES

whether humans live or die. It crosses new territory. Machines don't have our moral compass, our compassion and our emotions. Machines are not moral beings.

The technical argument is that these are potentially weapons of mass destruction, and the international community has thus far banned all other weapons of mass destruction.

What makes these different from previously banned weaponry is their potential to discriminate. You could say, "Only kill children," and then add facial recognition software to the system.

Moreover, if these weapons are produced, they would unbalance the world's geopolitics. Autonomous robotic weapons would be cheap and easy to produce. Some can be made with a 3-D printer, and they could easily fall into the hands of terrorists.

Another thing that makes them terribly destabilizing is that with such weapons, it

Toby Walsh, an authority in artificial intelligence at the University of New South Wales in Sydney, Australia. He is working with a coalition of scientists and human rights leaders seeking to halt the development of autonomous robotic weapons.

'You can't have machines deciding whether humans live or die.'

Continued



would be difficult to know the source of an attack. This has already happened in the current conflict in Syria. Just last year, there was a drone attack on a Russian-Syrian base, and we don't know who was actually behind it.

Why ban a weapon before it is produced?

The best time to ban such weapons is before they're available. It's much harder once they are falling into the wrong hands or becoming an accepted part of the military tool kit. The 1995 blinding laser treaty is perhaps the best example of a successful preemptive ban. Sadly, with almost every other weapon that has been regulated, we didn't have the foresight to do so in advance of it being used.

Your petition — whom was it addressed to?

The United Nations. Whenever I go there, people seem willing to hear from us. I never in my wildest dreams expected to be sitting down with the under secretary general of the U.N. and briefing him about the technology. One high U.N. official told me, "We rarely get scientists speaking with one voice. So when we do, we listen."

So far, 28 member countries have indicated their support. The European Parliament has called for it. The German foreign minister has called for it. Still, 28 countries out of 200! That's not a majority.

Who opposes the treaty?

The obvious candidates are the U.S., the U.K., Russia, Israel, South Korea. China has called for a pre-emptive ban on deployment, but not on development of the weapons.

It's worth pointing out there is going to be a huge amount of money being made by companies selling these weapons and the defenses to them.

Proponents of robotic weapons argue that by limiting the number of human combatants, the machines might make warfare less deadly.

I've heard those arguments, too. Some say that machines might be more ethical because people in warfare get frightened and do terrible things. Some supporters of the technology hope that this wouldn't happen if we had robots fighting wars, because they can be programmed to abide by interna-

tional humanitarian law.

The problem with that argument is that we don't have any way to program for something as subtle as international humanitarian law. Now, there are some things that the military can use robotics for — clearing a minefield is an example. If a robot goes in gets blown up, you get another robot.

Robotic warfare has long been the subject of science fiction films. Do you have a favorite?

No, most A.I. researchers — myself included — dislike how Hollywood has dealt with the technology. Kubrick's "2001" is way off because it is based on the idea that there will be machines that will have the desire for self-preservation, and that will result in malevolence toward humans.

It's wrong to assume they'll want to take over, or even preserve themselves. The intelligence we build is going to be quite different from what humans have, and it won't necessarily have the same character flaw.

These machines don't have any core science, and they don't have any desire to preserve themselves. They'll do exactly what we tell them to do. They are the most literal devices ever built.

I dislike "The Terminator," too. That technology is far, far away. There are more mundane technologies we should be worried about now, like the drones I mentioned.

Now, I do like "Her," because it is about the relationships we'll have in a future where we'll be increasingly interacting with machines. It will be possible, as in the movie that we will develop feelings for them.

That movie is about how A.I. is going to be a pervasive part of our existence in every room, every car. They will be things that listen to us, answer our questions and "understand" us.

Since 2013, you've been spending as much time on your activism as you have on scientific research. Any regrets?

No. This is important to be doing right now. Twenty years ago, like many of my colleagues, I felt that what we were doing A.I. was so far from practice that we didn't have to worry about moral consequences. That's no longer true. I have a 10-year-old daughter. When she's grown, I don't want her to ask, "Dad, you had a platform and authority — why didn't you try to stop this?"