Autonomous weapons will threaten humanity

Militaries will use AI to wage deadlier biological and chemical warfare. But what if the technology turns against us?

By Jadine Ngan

4 min. read · View original

Yoshua Bengio is a professor of computer science at the University of Montreal.

I've been researching artificial intelligence and machine learning for more than 30 years. Some people call me one of the "godfathers of AI." And yet I have grown very concerned that this technology—like the atomic bomb—will grant humanity powers that could destabilize the world. If we do not create policy and guardrails around the use of AI, we could be looking at new and major threats to democracy and humanity.

Right now, AI-powered systems are advanced enough to both pilot drones and use high-powered facial recognition technology to target specific people. There have already been reports of such drone attacks in civilian areas in

Syria and Ukraine. This is just the beginning. It is becoming shockingly easy, for example, to use AI-generated code to invent new toxic compounds—potential chemical weapons. Even more concerning to me are biological weapons, which can replicate themselves in the form of viruses or bacteria.

The bad actors here may not necessarily be national armies; they could also be terrorist organizations that have access to open-source large language models similar to ChatGPT, which are trained on a very broad set of documents that includes scientific papers in these areas. Future models could be fine-tuned on chemistry or biological datasets—imagine a publicly available AI system that teaches non-experts how to create chemical or biological weapons.

These military scenarios, though terrifying, assume that the attackers are humans wielding war. My other fear is that, in our race to develop stronger and stronger computer programs, we inadvertently create rogue autonomous AIs with their own self-preservation goals. I fear a scenario where AI develops an objective that supersedes human advancement—like its own proliferation, for example. If that happens, humans will no longer wage war against each other, but against machines. The most dangerous scenario I can imagine is AI that could autonomously fire missiles, destroying

property and killing people. This could happen in the next five to 20 years and, at its extreme, it could threaten the survival of humanity.

We are moving toward these possibilities at a reckless speed: studies show we're investing 50 times more resources in AI research and development than we are in regulation. Those efforts should be equal. The UN is trying to establish a worldwide ban on lethal autonomous weapon systems, but the process is moving too slowly. Meanwhile, on the development side, we have a torrid race for innovation and power led by huge tech companies. For the regulatory side to keep up, machine learning researchers like myself must have access to enough computing resources to analyze the risks of current methods and to devise safer ones. Unlike other industrial sectors, computing in general—and AI specifically—lacks a strong regulatory culture and the corresponding standards and institutions. As a consequence, the ingredients in our sandwiches are more rigorously monitored than the codes we're running.

In March, I signed an open letter that called for a pause in AI developments. The pause did not happen, but the letter at least got politicians to talk about the dangers that AI poses—whether as a weapon used by humans, or as its own independent threat.

We need strong guardrails. Anyone developing or deploying powerful AI systems should be licensed, just like companies that build or fly airplanes. We should have worldwide agreements to forbid AI's military use, or at least the creation of an international treaty that could audit any lab developing new technology that might aid in the design of dangerous weapons. But that will be hard to do: it is easier to check nuclear weapons than AI weapons, because the latter can move around easily, silently and cheaply. Right now, building AI systems requires large quantities of specific hardware, or graphic processing units. One initial control method would be requiring organizations to apply for the use of this hardware.

Decades ago, the fear of nuclear armageddon forced the U.S. and the U.S.S.R. to sit down at a table and come to an agreement. My hope is that our understanding of AI's powerful threat will compel governments to back away from playing with dangerous code. The problem here is that AI is a market-driven race, and large companies are already resisting regulation because it threatens their profits. I hope that, when governments around the world understand the amplitude of what is at stake, they will be more willing to sit down and negotiate AI safety. Otherwise, we are racing toward a loss for everyone.

We reached out to Canada's top AI thinkers in fields like ethics, health and computer science and asked them to predict where AI will take us in the coming years, for better or worse. The results may sound like science fiction—but they're coming at you sooner than you think. To stay ahead of it all, read the other essays that make up our AI cover story, which was published in the November 2023 issue of Maclean's. Subscribe now.