UN fails to agree on 'killer robot' ban as nations pour billions into autonomous weapons research

December 21 2021, by James Dawes

The Kargu-2, made by a Turkish defense contractor, is a cross between a quadcopter drone and a bomb. It has artificial intelligence for finding and tracking targets, and might have been used autonomously in the Libyan civil war to attack people. Credit: Ministry of Defense of Ukraine, CC BY 4.0

Autonomous weapon systems—commonly known as killer robots—may have killed human beings for the first time ever last year, according to a recent United Nations Security Council report on the Libyan civil war.

History could well identify this as the starting point of the next major arms race, one that has the potential to be humanity's final one.

The United Nations Convention on Certain Conventional Weapons debated the question of banning autonomous weapons at its once-every-five-years review meeting in Geneva Dec. 13–17, 2021, but didn't reach consensus on a ban. Established in 1983, the convention has been updated regularly to restrict some of the world's cruelest conventional weapons, including land mines, booby traps and incendiary weapons.

Autonomous weapon systems are robots with lethal weapons that can operate independently, selecting and attacking targets without a human weighing in on those decisions. Militaries around the world are investing heavily in autonomous weapons research and development. The U.S. alone budgeted US$18 billion for autonomous weapons between 2016 and 2020.

Meanwhile, human rights and humanitarian organizations are racing to establish regulations and prohibitions on such weapons development. Without such checks, foreign policy experts warn that disruptive autonomous weapons technologies will dangerously destabilize current nuclear strategies, both because they could radically change perceptions of strategic dominance, increasing the risk of preemptive attacks, and because they could be combined with chemical, biological, radiological and nuclear weapons themselves.

As a specialist in human rights with a focus on the weaponization of artificial intelligence, I find that autonomous weapons make the unsteady balances and fragmented safeguards of the nuclear world—for example, the U.S. president's minimally constrained authority to launch a strike—more unsteady and more fragmented. Given the pace of research and development in autonomous weapons, the U.N. meeting might have been the last chance to head off an arms race.

Lethal errors and black boxes

I see four primary dangers with autonomous weapons. The first is the problem of misidentification. When selecting a target, will autonomous weapons be able to distinguish between hostile soldiers and 12-year-olds playing with toy guns? Between civilians fleeing a conflict site and insurgents making a tactical retreat?

The problem here is not that machines will make such errors and humans won't. It's that the difference between human error and algorithmic error is like the difference between mailing a letter and tweeting. The scale, scope and speed of killer robot systems—ruled by one targeting algorithm, deployed across an entire continent—could make misidentifications by individual humans like a recent U.S. drone strike in Afghanistan seem like mere rounding errors by comparison.

Autonomous weapons expert Paul Scharre uses the metaphor of the runaway gun to explain the difference. A runaway gun is a defective machine gun that continues to fire after a trigger is released. The gun continues to fire until ammunition is depleted because, so to speak, the gun does not know it is making an error. Runaway guns are extremely dangerous, but fortunately they have human operators who can break the ammunition link or try to point the weapon in a safe direction. Autonomous weapons, by definition, have no such safeguard.

Importantly, weaponized AI need not even be defective to produce the runaway gun effect. As multiple studies on algorithmic errors across industries have shown, the very best algorithms—operating as designed—can generate internally correct outcomes that nonetheless spread terrible errors rapidly across populations.

For example, a neural net designed for use in Pittsburgh hospitals identified asthma as a risk-reducer in pneumonia cases; image

recognition software used by Google identified Black people as gorillas; and a machine-learning tool used by Amazon to rank job candidates systematically assigned negative scores to women.

The problem is not just that when AI systems err, they err in bulk. It is that when they err, their makers often don't know why they did and, therefore, how to correct them. The black box problem of AI makes it almost impossible to imagine morally responsible development of autonomous weapons systems.

The proliferation problems

The next two dangers are the problems of low-end and high-end proliferation. Let's start with the low end. The militaries developing autonomous weapons now are proceeding on the assumption that they will be able to contain and control the use of autonomous weapons. But if the history of weapons technology has taught the world anything, it's this: Weapons spread.

Market pressures could result in the creation and widespread sale of what can be thought of as the autonomous weapon equivalent of the Kalashnikov assault rifle: killer robots that are cheap, effective and almost impossible to contain as they circulate around the globe. "Kalashnikov" autonomous weapons could get into the hands of people outside of government control, including international and domestic terrorists.

High-end proliferation is just as bad, however. Nations could compete to develop increasingly devastating versions of autonomous weapons, including ones capable of mounting chemical, biological, radiological and nuclear arms. The moral dangers of escalating weapon lethality would be amplified by escalating weapon use.

High-end autonomous weapons are likely to lead to more frequent wars because they will decrease two of the primary forces that have historically prevented and shortened wars: concern for civilians abroad and concern for one's own soldiers. The weapons are likely to be equipped with expensive ethical governors designed to minimize collateral damage, using what U.N. Special Rapporteur Agnes Callamard has called the "myth of a surgical strike" to quell moral protests. Autonomous weapons will also reduce both the need for and risk to one's own soldiers, dramatically altering the cost-benefit analysis that nations undergo while launching and maintaining wars.

Asymmetric wars—that is, wars waged on the soil of nations that lack competing technology—are likely to become more common. Think about the global instability caused by Soviet and U.S. military interventions during the Cold War, from the first proxy war to the blowback experienced around the world today. Multiply that by every country currently aiming for high-end autonomous weapons.

Undermining the laws of war

Finally, autonomous weapons will undermine humanity's final stopgap against war crimes and atrocities: the international laws of war. These laws, codified in treaties reaching as far back as the 1864 Geneva Convention, are the international thin blue line separating war with honor from massacre. They are premised on the idea that people can be held accountable for their actions even during wartime, that the right to kill other soldiers during combat does not give the right to murder civilians. A prominent example of someone held to account is Slobodan Milosevic, former president of the Federal Republic of Yugoslavia, who was indicted on charges of crimes against humanity and war crimes by the U.N."s International Criminal Tribunal for the Former Yugoslavia.

But how can autonomous weapons be held accountable? Who is to blame

for a robot that commits war crimes? Who would be put on trial? The weapon? The soldier? The soldier's commanders? The corporation that made the weapon? Nongovernmental organizations and experts in international law worry that autonomous weapons will lead to a serious accountability gap.

To hold a soldier criminally responsible for deploying an autonomous weapon that commits war crimes, prosecutors would need to prove both actus reus and mens rea, Latin terms describing a guilty act and a guilty mind. This would be difficult as a matter of law, and possibly unjust as a matter of morality, given that autonomous weapons are inherently unpredictable. I believe the distance separating the soldier from the independent decisions made by autonomous weapons in rapidly evolving environments is simply too great.

The legal and moral challenge is not made easier by shifting the blame up the chain of command or back to the site of production. In a world without regulations that mandate meaningful human control of autonomous weapons, there will be war crimes with no war criminals to hold accountable. The structure of the laws of war, along with their deterrent value, will be significantly weakened.

A new global arms race

Imagine a world in which militaries, insurgent groups and international and domestic terrorists can deploy theoretically unlimited lethal force at theoretically zero risk at times and places of their choosing, with no resulting legal accountability. It is a world where the sort of unavoidable algorithmic errors that plague even tech giants like Amazon and Google can now lead to the elimination of whole cities.

In my view, the world should not repeat the catastrophic mistakes of the nuclear arms race. It should not sleepwalk into dystopia.

Provided by The Conversation