We Should Not Ban ‘Killer Robots,’ and Here’s Why

What we really need is a way of making autonomous armed robots ethical, because we’re not going to be able to prevent them from existing.

By Evan Ackerman

Yesterday, an open letter was presented at the International Joint Conference on Artificial Intelligence in Buenos Aires, Argentina, calling for a “ban on offensive autonomous weapons.” A bunch of people signed it, including “more than 1,000 experts and leading robotics researchers.” And I mean, of course they’d sign it, because who would seriously be for “killer robots?”

I am.

Here’s the letter in full:

Autonomous Weapons: an Open Letter from AI & Robotics Researchers

Autonomous weapons select and engage targets without human intervention. They might include, for example, armed quadcopters that can search for and eliminate people meeting certain pre-defined criteria, but do not include cruise missiles or remotely piloted drones for which humans make all targeting decisions. Artificial Intelligence (AI) technology has reached a point where the deployment of such systems is — practically if not legally — feasible within years, not decades, and the stakes are high: autonomous weapons have been described as the third revolution in warfare, after gunpowder and nuclear arms.
Many arguments have been made for and against autonomous weapons, for example that replacing human soldiers by machines is good by reducing casualties for the owner but bad by thereby lowering the threshold for going to battle. The key question for humanity today is whether to start a global AI arms race or to prevent it from starting. If any major military power pushes ahead with AI weapon development, a global arms race is virtually inevitable, and the endpoint of this technological trajectory is obvious: autonomous weapons will become the Kalashnikovs of tomorrow. Unlike nuclear weapons, they require no costly or hard-to-obtain raw materials, so they will become ubiquitous and cheap for all significant military powers to mass-produce. It will only be a matter of time until they appear on the black market and in the hands of terrorists, dictators wishing to better control their populace, warlords wishing to perpetrate ethnic cleansing, etc. Autonomous weapons are ideal for tasks such as assassinations, destabilizing nations, subduing populations and selectively killing a particular ethnic group. We therefore believe that a military AI arms race would not be beneficial for humanity. There are many ways in which AI can make battlefields safer for humans, especially civilians, without creating new tools for killing people.

Just as most chemists and biologists have no interest in building chemical or biological weapons, most AI researchers have no interest in building AI weapons — and do not want others to tarnish their field by doing so, potentially creating a major public backlash against AI that curtails its future societal benefits. Indeed, chemists and biologists have broadly supported international agreements that have successfully prohibited chemical and biological weapons, just as most physicists supported the treaties banning space-based nuclear weapons and blinding laser weapons.

In summary, we believe that AI has great potential to benefit humanity in many ways, and that the goal of the field should be to do so. Starting a military AI arms race is a bad idea, and should be prevented by a ban on offensive autonomous weapons beyond meaningful human control.

The main point in the body of this letter seems to be that unless we outlaw autonomous weapons right now, there will be some sort of arms race that will lead to the rapid advancement and propagation of things like autonomous “armed quadcopters,” eventually resulting in technology that’s accessible to anyone if they want to build a weaponized drone.
The problem with this argument is that no letter, UN declaration, or even a formal ban ratified by multiple nations is going to prevent people from being able to build autonomous, weaponized robots. The barriers keeping people from developing this kind of system are just too low. Consider the “armed quadcopters.”

Today you can buy a smartphone-controlled quadrotor for US $300 at Toys R Us. Just imagine what you’ll be able to buy tomorrow. This technology exists. It’s improving all the time. There’s simply too much commercial value in creating quadcopters (and other robots) that have longer endurance, more autonomy, bigger payloads, and everything else that you’d also want in a military system. And at this point, it’s entirely possible that small commercial quadcopters are just as advanced as (and way cheaper than) small military quadcopters, anyway. We’re not going to stop that research, though, because everybody wants delivery drones (among other things). Generally speaking, technology itself is not inherently good or bad: it’s what we choose to do with it that’s good or bad, and you can’t just cover your eyes and start screaming “STOP!!!” if you see something sinister on the horizon when there’s so much simultaneous potential for positive progress.

What we really need, then, is a way of making autonomous armed robots ethical, because we’re not going to be able to prevent them from existing. In fact, the most significant assumption that this letter makes is that armed autonomous robots are inherently more likely to cause unintended destruction and death than armed autonomous humans are. This may or may not be the case right now, and either way, I genuinely believe that it won’t be the case in the future, perhaps the very near future. I think that it will be possible for robots to be as good (or better) at identifying hostile enemy combatants as humans, since there are rules that can be followed (called Rules of Engagement, for an example see page 27 of this) to determine whether or not using force is justified. For example, does your target have a weapon? Is that weapon pointed at you? Has the weapon been fired? Have you been hit? These are all things that a robot can determine using any number of sensors that currently exist.

It’s worth noting that Rules of Engagement generally allow for engagement in the event of an imminent attack. In other words, if a hostile target has a weapon and that weapon is pointed at you, you can engage before the weapon is fired rather than after in the interests of self-protection. Robots could be even more cautious than this: you could program them to not engage a hostile target with deadly force unless they confirm with whatever level of certainty that you want that the target is actively engaging them already. Since robots aren’t alive and don’t have emotions and don’t get tired or stressed or distracted, it’s possible for them to just sit there, under fire, until all necessary criteria for engagement are met. Humans can’t do this.

“What we really need is a way of making autonomous armed robots ethical, because we’re not going to be able to prevent them from existing”
The argument against this is that a robot autonomously making a decision to engage a target with deadly force, no matter how certain the robot may be, is dangerous and unethical. It is dangerous, and it may be unethical, as well. However, is it any more dangerous or unethical than asking a human to do the same thing? The real question that we should be asking is this: Could autonomous armed robots perform better than armed humans in combat, resulting in fewer casualties (combatant or non-combatant) on both sides? I believe so, which doesn’t really matter, but so do people who are actually working on this stuff, which does.

In 2009, Ronald C. Arkin, Patrick Ulam, and Brittany Duncan published a paper entitled “An Ethical Governor for Constraining Lethal Action in an Autonomous System,” which was about how to program an armed, autonomous robot to act within the Laws of War and Rules of Engagement. h+ Magazine interviewed Arkin on the subject (read the whole thing here), and here’s what he said:

**h+:** Some researchers assert that no robots or AI systems will be able to discriminate between a combatant and an innocent, that this sensing ability currently just does not exist. Do you think this is just a short-term technology limitation? What such technological assumptions do you make in the design of your ethical governor?

**RA:** I agree this discrimination technology does not effectively exist today, nor is it intended that these systems should be fielded in current conflicts. These are for the so-called war after next, and the DoD would need to conduct extensive additional research in order to develop the accompanying technology to support the proof-of-concept work I have developed. But I don’t believe there is any fundamental scientific limitation to achieving the goal of these machines being able to discriminate better than humans can in the fog of war, again in tightly specified situations. This is the benchmark that I use, rather than perfection. But if that standard is achieved, it can succeed in reducing noncombatant casualties and thus is a goal worth pursuing in my estimation.

One way to think about this is like autonomous cars. Expecting an autonomous car to keep you safe 100 percent of the time is unrealistic. But, if an autonomous car is (say) 5 percent more likely to keep you safe than if you were driving yourself, you’d still be much better off letting it take over. Autonomous cars, by the way, will likely be much safer than that, and it’s entirely possible that autonomous armed robots will be, too. And if autonomous armed robots really do have at least the potential reduce casualties, aren’t we then ethically obligated to develop them?
If there are any doubts about how effective or ethical these systems might be, just test them exhaustively. Deploy them, load them up with blanks, and watch how they do. Will they screw up sometimes? Of course they will, both during testing and after. But setting aside the point above about relative effectiveness, the big advantage of robots is that their behavior is traceable and they learn programmatically: if one robot does something wrong, it’s possible to trace the chain of decisions that it made (decisions programmed into it by a human, by the way) to find out what happened. Once the error is located, it can be resolved, and you can be confident that the robot will not make that same mistake again. Furthermore, you can update every other robot at the same time. This is not something we can do with humans.

I do agree that there is a potential risk with autonomous weapons of making it easier to decide to use force. But, that’s been true ever since someone realized that they could throw a rock at someone else instead of walking up and punching them. There’s been continual development of technologies that allow us to engage our enemies while minimizing our own risk, and what with the ballistic and cruise missiles that we’ve had for the last half century, we’ve got that pretty well figured out. If you want to argue that autonomous drones or armed ground robots will lower the bar even farther, then okay, but it’s a pretty low bar as is. And fundamentally, you’re then placing the blame on technology, not the people deciding how to use the technology.

“I’m not in favor of robots killing people. If this letter was about that, I’d totally sign it. But that’s not what it’s about; it’s about the potential value of armed autonomous robots, and I believe that this is something that we need to have a reasoned discussion about rather than banning.”

And that’s the point that I keep coming back to on this: blaming technology for the decisions that we make involving it is at best counterproductive and at worst nonsensical. Any technology can be used for evil, and many technologies that were developed to kill people are now responsible for some of our greatest achievements, from harnessing nuclear power to riding a ballistic missile into space. If you want to make the argument that this is really about the decision to use the technology, not the technology itself, then that’s awesome. I’m totally with you. But banning the technology is not going to solve the problem if the problem is the willingness of humans to use technology for evil: we’d need a much bigger petition for that.
I want to be very clear about this: I’m not in favor of robots killing people. If this letter was about that, I’d totally sign it. But that’s not what it’s about; it’s about the potential value of armed autonomous robots, and I believe that this is something that we need to have a reasoned discussion about rather than banning. I’m open to the fact that I might be quite wrong about every point that I’ve made here, but the important thing is to be able to reach an informed decision rather than just demanding to outlaw “killer robots.”*

* I hate the term “killer robots,” and I hate that it was used by the authors of the letter to promote it. Roboticists have spent a long, long time fighting against this killer robots trope, which has been relentlessly propagated through popular culture. And now, a bunch of smart people who really should know better are using that same "killer robots" rhetoric to try and frighten people into agreeing with them. Does anyone want more “killer robots?” Of course not. Is it actually that simple? Of course it isn’t. But, if you instead call for reasoned debate about armed autonomous systems, you’re not going to get as many emotional reactions in your favor.

Robotics News

Biweekly newsletter on advances and news in robotics, automation, control systems, interviews with leading roboticists, and more.

About the Automaton blog

IEEE Spectrum's award-winning robotics blog, featuring news, articles, and videos on robots, humanoids, automation, artificial intelligence, and more.

Erico Guizzo, Editor

Evan Ackerman, Senior Writer

Fan Shi, Contributor

Subscribe to RSS Feed