

Much Ado About Killer Robots

Written by Alexander Leveringhaus

This PDF is auto-generated for reference only. As such, it may contain some conversion errors and/or missing information. For all formal use please refer to the official version on the website, as linked below.

Much Ado About Killer Robots

<https://www.e-ir.info/2013/06/04/much-ado-about-killer-robots/>

ALEXANDER LEVERINGHAUS, JUN 4 2013

The March of the Killer Robot

'A spectre is haunting the battlefields of the world. And the combined powers of the UN and civil society have united in a holy alliance to exorcise it'. The spectre in question is the Killer Robot. Mercilessly, the Killer Robot marches across the battlefield, mowing down everything in its path. Indeed, we are, Human Rights Watch warns us, on the verge of losing humanity in war by conveniently outsourcing the dirty work of killing to machines.[i] But regrettably, machines neither have compassion nor common sense. Rather, like the Terminator (minus the Austrian accent), they'll blow up everything in their way. Faced with the prospect of robo-wars, even the UN Special Rapporteur on Arbitrary Executions and Extra-Judicial Killings, Christof Heyns, has called for a moratorium on the development of Killer Robots, while Human Rights Watch wants to see them banned.[ii]

The only problem is that it is by no means clear what Killer Robots are. Some of the relevant systems, as Human Rights Watch admits, don't even exist yet. This leads to a curious situation: how can one ban something without knowing precisely what one is banning? But perhaps the current debate is less a matter of precise definitions but gut feeling. Something has gone wrong somewhere: operators of unmanned aerial vehicles (also known as drones) can now target individuals at the push of a button without ever having to leave their cubicle which is usually located thousands of miles away from the actual battlefield. According to critics, war becomes a computer game, played, as Christopher Coker puts it, by Geeks rather than Greeks.[iii] The next step consists in taking the human out of the decision-making loop entirely. Machines may be much more efficient in disposing of our enemies than we are. But then again, humans have not needed much persuasion to kill each other, be it in the name of religion, the nation, the tribe, the state, or class. Are robots really any worse? So, what is the fuss all about?

Robots Doing it Themselves

Human Rights Watch and many of their (academic) supporters are worried about robotic weapons with a high degree of what I call operational autonomy. The latter term simply denotes that a machine can carry out a particular task without assistance from an operator. Accordingly, an operationally autonomous weapon can sense, identify, track and destroy a target all by itself. In this sense, it differs from a remote-controlled weapon, such as an unmanned aerial vehicle, where targeting decisions are made by a human operator via remote control. It also differs from systems that are operationally autonomous with regard to some tasks, e.g. navigating a complex environment by themselves, but lack operational autonomy with regard to the application of lethal force to a target. A Killer Robot, then, is a robotic weapon that is operationally autonomous with regard to targeting. That said, not every autonomous weapon is necessarily a robot. Some systems may be robotic; others may not be robots at all or merely contain some robotic elements. But these definitional issues don't need to concern us here. For the sake of convenience, I use 'Killer Robot' as an umbrella term for autonomous weapons in general.[iv]

Now, it is noteworthy that autonomous weapons already exist. A landmine, for instance, could be seen as an operationally autonomous weapon. It can blow up anyone who steps on it without an operator. But perhaps, the difference between landmines and Killer Robots is that the former are mere mechanical devices, while the latter are cognitive systems. (By the way, anyone reading this is a cognitive system, too.) That is, Killer Robots can (1) acquire information about their environment via sensors, (2) analyse that information, (3) decide how to proceed, and (4)

Much Ado About Killer Robots

Written by Alexander Leveringhaus

enact the decision. But this is nothing new. Modern missile defence systems, for instance, are cognitive systems, though they are usually not fully operationally autonomous. Traditionally, they are operationally autonomous with regard to (1) and (4) but not (2) and (3). But if we already have systems that *a*) are cognitive in nature and *b*) operationally autonomous, albeit to varying degrees, why should we be worried about Killer Robots *now*? Any attempt to answer this question reveals, I believe, three fallacies in the Killer Robots debate.

Three Fallacies in the Killer Robots Debate:

The military will use Killer Robots in every context and for every conceivable task:

It seems that Killer Robots, for many critics, are analogous to human combatants. That is, just as human combatants are ordered to fight against other human combatants, Killer Robots will be ordered to fight against human combatants. One worry in this regard is that Killer Robots cannot comply with the laws of war – especially the principle of discrimination – because they cannot distinguish between combatants and non-combatants. It is already hard enough for humans to apply this distinction, but due to the complexity of human behaviour, this is, critics contend, nearly impossible for machines. The critics are right, of course. But many (though not all) operationally autonomous systems that are being developed are not specifically designed to target humans (or fight against them). Rather, they are engineered to destroy targets that, unlike humans, have a (relatively) unambiguous ‘signature’. For instance, the Taranis stealth aircraft, currently being developed by BAE Systems, tracks the signals emitted by radar stations and destroys them without the assistance of an operator. It is not clear, then, that Killer Robots will operate on a battlefield in the same way as humans. Machines will take over *some* tasks in war, but I don’t think that even the staunchest defenders of these weapons claim that they will be suitable for *all* tasks.^[v] The key question is whether, in a *particular* context, it would be legal to use a Killer Robot for a *specific* task. Sometimes the answer may be positive, sometimes it is likely to be negative. A blanket ban, at first sight, is unjustified. Coincidentally, we should abandon talk of robo-armies. It is true that, in some areas of the military, Killer Robots will be used to complement or even replace humans. But in other areas their use will be limited.

War becomes riskless, never mind Russian nukes:

Some critics of Killer Robots claim that we are going to see more wars because reliance on these weapons renders war riskless. Similar claims are often made about (non-autonomous) remote-controlled weapons. But they are hard to believe. Taking US drone policy as an example, critics are right to point out that drone attacks in Pakistan and Yemen have greatly increased. But it is questionable whether this reveals a general point about the alleged riskless nature of war. While drone attacks in Pakistan and Yemen have increased, the USA, despite her impressive arsenal of drones, has been anxious (at the time of writing) to avoid intervening in the civil war in Syria. If drones really made it entirely riskless to take out the Syrian president, why is no one pressing the button? The answer is that technology in and of itself may enhance the military capacities of states. But it is not the only element which decides whether states go to war. The US government may deem it necessary to hunt down (alleged) members of al-Qaeda, but it is not keen on a confrontation with China and Russia, the main supporters of the Syrian government. Nor does it have any interest in long-term nation building in Syria, or in the creation a power vacuum that might be exploited by Islamic fundamentalists. In general, in a world where powerful states have a frightening arsenal of weapons of mass destruction, war is an inherently risky activity. Let a Killer Robot destroy a target in Russia and see what Mr. Putin has to say, provided you live long enough to hear his response, of course.

It is morally perverse that Killer Robots are going to decide about life and death:

Critics claim that it is morally perverse that Killer Robots should ‘choose’ their targets. Machines, the argument goes, should not make decisions about life and death. There are variations of this criticism. But it is not necessary to go into detail here, for, I believe, they are all mistaken for two reasons.

First, they all seem to assume that a Killer Robot is a moral agent. What constitutes moral agency is hotly disputed in moral philosophy and metaphysics.^[vi] To be sure, a Killer Robot in virtue of being a cognitive system is an agent because it can interact with its environment and adjust its behaviour accordingly. But this is not sufficient to qualify as

Much Ado About Killer Robots

Written by Alexander Leveringhaus

a *moral* agent. For that to be the case we would have to be justified in adopting what the philosopher Peter Strawson famously called 'reactive attitudes' to the robot by praising or blaming it for its actions.[vii] But we are not justified in doing so: the robot is a device specifically engineered by humans to carry out particular tasks. It is not a moral agent in its own right. The question, as I indicated earlier, is whether it is morally permissible and legal for *humans* to use a Killer Robot in order to destroy certain targets. The question whether *robots* should be *morally* permitted to kill humans is irrelevant.

Second, the whole point of having a military is that you don't choose your own targets. Your targets are chosen for you. In other words, you will be acting under orders. The same is true of Killer Robots. They will be programmed by humans to attack particular targets. Critics of Killer Robots sound as if the military will let machines decide whom to kill. If that was the case, the military would, in effect, abolish itself as an institution. Budget cuts and austerity are one thing, but I find it hard to believe that the military is keen on getting rid of itself.

All Plain Sailing?

I have tried to debunk some of the arguments levelled against Killer Robots. But does this mean that there is nothing to worry about? Alas, academics are notorious pessimists, and I am no exception. No, we should not relax, for there remain many unanswered questions. First and foremost, it would have to be clarified in which contexts it would be legal to deploy a Killer Robot. Moreover, we also need to ask how great the risk of 'hacking' or 're-programming' by enemies would be. Finally, we must find out what the worst case scenario would be if the algorithms of different Killer Robots began to interact with each other.[viii] Are these weapons safe to operate, or are they inherently uncontrollable? There are surely many more questions. Let us have a proper debate about them.

—

Alex Leveringhaus is a post-doctoral researcher based at the Oxford Institute for Ethics, Law and Armed Conflict (ELAC), where he works on the Military Human Enhancement Project run in cooperation between Delft University of Technology and ELAC. At Oxford, he is also a James Martin Fellow in the Oxford Martin School and a Research Member of Wolfson College Oxford. Alex trained as a political philosopher at the LSE and has published on humanitarian intervention as well as new military technologies: 'The Moral Status of Combatants during Military Humanitarian Intervention', *Utilitas*, 24/2 (2012) & (with Tjerk de Greef) 'Tele-operated Weapons Systems: Safeguarding Moral Perception and Responsibility', in M. Aaronson & A. Johnson (eds.), *Hitting the Target: How New Capabilities are Shaping International Intervention* (London: Royal United Services Institute, 2013).

[i] Human Rights Watch, *Losing Humanity: The Case against Killer Robots*, available at <http://www.hrw.org/reports/2012/11/19/losing-humanity-0>, accessed 03/06/2013.

[ii] Christof Heyns, Report of the Special Rapporteur on extra-judicial, summary and arbitrary executions, available at <http://daccess-dds-ny.un.org/doc/UNDOC/GEN/G13/127/76/PDF/G1312776.pdf?OpenElement>.

[iii] C. Coker, *From Greeks to Geeks* (London: Hurst, 2013).

[iv] One interesting question is whether the fact that a specific autonomous weapon is a robot gives rise to any distinctive legal and moral issues. I am not sure it does.

[v] I'd like to thank Ron Arkin for discussions about this.

[vi] Mark Rowlands offers some thought-provoking views on what constitutes moral agency in his

Much Ado About Killer Robots

Written by Alexander Leveringhaus

Can Animals Be Moral? (Oxford: Oxford University press, 2013).

[vii] P. Strawson, 'Freedom and Resentment', reprinted in R. Shafer-Landau (ed.), *Ethical Theory: An Anthology*, 2nd ed., (Oxford: Oxford University Press, 2013).

[viii] I am grateful to Noel Sharkey for alerting me to this point.

About the author:

Alex Leveringhaus is a post-doctoral researcher based at the Oxford Institute for Ethics, Law and Armed Conflict (ELAC), where he works on the Military Human Enhancement Project run in cooperation between Delft University of Technology and ELAC. At Oxford, he is also a James Martin Fellow in the Oxford Martin School and a Research Member of Wolfson College Oxford. Alex trained as a political philosopher at the LSE and has published on humanitarian intervention as well as new military technologies: 'The Moral Status of Combatants during Military Humanitarian Intervention', *Utilitas*, 24/2 (2012) & (with Tjerk de Greef) 'Tele-operated Weapons Systems: Safeguarding Moral Perception and Responsibility', in M. Aaronson & A. Johnson (eds.), *Hitting the Target: How New Capabilities are shaping international intervention* (London: Royal United Services Institute, 2013).