

# Jury Still Out on Killer Robots

Written by Paul Robinson

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.

## Jury Still Out on Killer Robots

<https://www.e-ir.info/2012/12/10/jury-still-out-on-killer-robots/>

PAUL ROBINSON, DEC 10 2012

Killer robots have long been a staple of science fiction. From the Doomsday Machine in the original Star Trek series, through Terminator and Battlestar Galactica, the idea of lethal machines running amok and destroying mankind both fascinates and terrifies. Until recently, such images have seemed like far-fetched fantasies of the distant future. Advances in technology mean, however, that the prospect of so-called 'lethal autonomous robots' is now far more real, and as a result pressure is growing to outlaw them before they become an established part of the international scene.

The most obvious manifestation of the new anti-robot movement is a recent report by Human Rights Watch entitled 'Losing Humanity: The Case Against Killer Robots' which calls for an international convention regulating the use of lethal autonomous robots. Human Rights Watch's report is far from being the first examination of the subject. Armin Krishnan of the University of Texas and Ronald Arkin of the Georgia Institute of Technology have studied the matter in more depth. Nevertheless, until now the subject of the legality and ethics of lethal autonomous robots has been rather an esoteric one which has not generated a great deal of publicity. Human Rights Watch have changed this, and for that reason the report is to be welcomed. But the 'case against killer robots' is probably not as clear cut as Human Rights Watch makes out.

The report makes three main arguments for an international convention outlawing lethal autonomous robots, that is to say machines which are capable of killing without any human involvement in the decision-making process. The first argument is one of accountability: a machine is neither a moral nor a legal agent, and so cannot be held to account if it commits what for a human being would be war crimes. The second argument is a practical one: machines are not capable of emotion and so would be unable to make correct ethical judgements. The third argument is broader: taking human beings out of war reduces the risks for policy makers and so encourages them to make more use of force – the result of the use of killer robots, in other words, might be more war.

Human Rights Watch is about half right. Before showing that, it is worth a short diversion to determine whether there is an issue here worth discussing. There are at least a couple of reasons for supposing that perhaps there is not.

First, the image of killer robots brings to mind Terminator-style machines which are not remotely on the horizon. Even with rapid technological progress, for the foreseeable future robots will remain fairly simple devices restricted to narrowly bounded operations. Within those narrow boundaries they may be very capable, but a robot able to replace a human being and operate fully independently remains pure science fiction. Any discussion of lethal autonomous robots must therefore take into consideration the limitations of what such machines will actually be capable of and the very restricted roles in which they will operate. Viewed this way, they become rather less scary.

This draws our attention to another point, which is that of definitions. What is a robot? And what do we mean by autonomous? The phrase 'lethal autonomous robot' suggests something fairly sophisticated. Yet, some people would deem a toaster to be a robot, even an autonomous one, and one could make a good argument that an old-fashioned landmine is a lethal autonomous robot, as it has its own sensor and using that, makes the decision to kill by itself without any human involvement once it has been turned on. Again this reinforces the point that lethal autonomous robots are rather more limited than people tend to think and there are gradations of 'autonomy' that make a legal definition for arms control decidedly difficult.

## Jury Still Out on Killer Robots

Written by Paul Robinson

Second, some critics argue that the entire robot/human dichotomy is a false one. In a recent article for *Slate* magazine, for instance, Brad Allenby of Arizona State University argued that this dichotomy 'is just too flawed and oversimplistic a foundation on which to build policy.' The future, Allenby suggests, lies in a merging of human and machines through techniques such as neural networking. The killer robots of the future will not be entirely autonomous but rather will be naturally connected to human operators. That possibility raises a whole host of ethical and legal questions, but not those which Human Rights Watch sees as important.

We are not yet in a position to know. It could be that the issue of lethal autonomous robots proves to be far less important than it seems at first sight. Still, it seems a little unwise to dismiss it straight out of hand. However we define robot or autonomy, technology is progressing in this area, and we should consider the implications, while bearing in mind the limitations mentioned above.

With this in mind, we can now consider whether Human Rights Watch's three arguments stand up to close scrutiny.

Of the three, the first, that of accountability, is the strongest. That said, it depends on the chosen definition of autonomous. There is no accountability problem with, say, a land mine, because, although it acts on its own, it is so unsophisticated that blame clearly lies with the person who deployed it. A truly autonomous robot, by contrast, one that could make decisions entirely by itself, would be responsible for its own actions. It would not be fair to blame the designer or the person who deployed the weapon for a decision made by the machine. But one cannot punish a machine – it cannot suffer in any meaningful way – and there would therefore be nobody to hold to account should the robot make a mistake or, even worse, deliberately break the laws of war. Giving the power to kill to objects which are not accountable is not a desirable state of affairs, and this therefore is a matter of serious concern. Whether this is serious enough to warrant outlawing lethal autonomous robots depends in part on how far along the spectrum from the stupid landmine to the truly intelligent autonomous robot we actually are.

The second argument, concerning machines' abilities to make ethical decisions, is a little shakier. Certainly emotion is an important part of ethical decision making. It is also true that ethical decisions are difficult and can involve a vast number of variables, especially in an environment as complex as war. It seems unlikely that a computer, at least one we can reasonably envisage in the near future, could cope with such complexity. That said, emotions such as fear and anger are a cause of misbehaviour, and the actual situations in which lethal autonomous robots are in practice likely to be deployed are, as I previously mentioned, rather narrowly bounded. In these circumstances it is not impossible that they could actually make better decisions than humans. The jury is out on this one, and while Human Rights Watch may eventually be proven right, it is far too early to rush to judgement.

Finally, the third argument, that autonomous robots will make it easier to use lethal force, misses the mark. Riskless war does have its dangers: if political leaders feel that they can use force without a great risk of their own people being killed, then indeed they may be more inclined to do so. But this has nothing to do with autonomous robots. It applies equally to non-autonomous ones. A drone operated by a human in Las Vegas involves no more real risk than an autonomous drone. Banning lethal autonomous robots will do nothing to solve this problem.

Overall, therefore, Human Rights Watch scores one and half out of three. This may be sufficient to make its case, but it is hardly a convincing outcome. One cannot say that the case for outlawing killer robots has been proven wrong, but it is certainly premature to say that it has been proven right.

—

**Paul Robinson** is a professor in the Graduate School of Public and International Affairs at the University of Ottawa, and the author and editor of numerous works on military history and military ethics.

## **Jury Still Out on Killer Robots**

Written by Paul Robinson

### **About the author:**

Paul Robinson is a Professor in the Graduate School of Public and International Affairs at the University of Ottawa. He is the author of numerous works on Russian and Soviet history, including most recently *Grand Duke Nikolai Nikolaevich: Supreme Commander of the Russian Army* (Northern Illinois University Press, 2014). He blogs at Irrussianality.