

Informal Meeting of Experts on Lethal Autonomous Weapons: Convention on Conventional Weapons

Geneva: 16 April 2016

Panel on Human Rights and Lethal Autonomous Weapons Systems (LAWS)

Comments by Christof Heyns, United Nations Special Rapporteur on extrajudicial, summary or arbitrary executions (as finalised after the meeting)

Madam Chair, excellences, ladies and gentlemen

Let me first of all commend the CCW on having arranged this meeting and this panel, and thank you for the invitation.

By way of **introduction**: to the chair mentioned my report on the issue of autonomous weapons in a May 2013 report to the Human Rights Council, which was of course preceded by a report to the General Assembly by my predecessor as mandate holder on extrajudicial executions, Philip Alston. The day before I delivered my report I spoke to a well-seasoned diplomat, who told me to tone down my expectations about a quick uptake on the issue by the UN. He told me such matters, important as they may be, take years to be gain attention, and even then it is very rare especially for something that has not yet caused major problems to be addressed in a pre-emptive way.

Looking back almost two years later I can say that I have been impressed by how quickly the issue was in fact taken up, and by the creativity and vigour with which a whole range of bodies within the UN as well as those outside, such as the ICRC and CSOs, have engaged with the matter. The role of the CCW in particular has been crucial. Especially significant is what appears to be an emerging consensus that the notion of meaningful human control presents a guide to distinguish acceptable forms of autonomous force release.

The emergence of this broad consensus constitutes to my mind a significant step towards a solution, and it illustrates how the dialogue between states can benefit

from being open to inputs from outsiders. In truth, the fact that the issue of autonomous weapons is on the agenda at all also testifies to this – it was after all a Human Rights Watch/Harvard Human Rights Programme report which really drew the international attention to this emerging issue.

It is still too early to say whether the second point of the diplomat's caution will be addressed – whether the talks will lead to concrete action. Let us recognise that so far much has been achieved, but much remains to be done.

I was a pleasure to attend many of the meetings held on autonomous weapons here in Geneva and in other parts of the world over the last two years. From my perspective it has been particularly useful that almost all the meetings – and this week is no exception – have been looking at this multi-faceted issue from a multi-disciplinary perspective, including the human rights perspective. It seems fair to say that there is much greater awareness today of the need to keep human rights standards in mind when dealing with use of force issues in general, including the use of force during armed conflict. That in itself is already a very positive and necessary development.

I would like to address four issues:

1. It may in the first place be appropriate to take a step back, and to **ask ourselves how we should articulate the underlying issue on the table** today. A popular approach may be to say it is concern about machines taking decisions about the use of lethal force against human beings during armed conflict. However, in my view autonomous weapons raise a larger concern. I think the broader concern is about machines taking decisions about the use of force against human beings, whether it is in armed conflict or not, and whether the force is lethal or not.

While the question of machines determining whether people will live or die during war is for current purposes the main issue to be addressed, it is not the only one. Increasingly law enforcement officials use unmanned systems, for example in riot control, and in some of these systems force may be released autonomously. Existing systems can for example deploy teargas without human intervention – and in future they may well be able to activate rubber bullets, tazers etc. These weapons are in the ordinary course of events not lethal, but the issues raised by such expansion of autonomous force release are in many respects the same as those raised by its use on the battlefield: essentially the concern is about robots having physical control over our lives.

We will be doing ourselves a disservice by dealing with one aspects of this problem in isolation from the other aspects, or if the disarmament and human rights communities go in different directions about the underlying issue. We may find that we end up with an incoherent set of terms and norms that try to address different parts of the same issue.

Several consequences follow if one is to take this more holistic approach:

- In the first place it seems that the better general term to be used is probably Autonomous Weapons Systems (AWS). LAWS is a sub-category of AWS, just as Fully Autonomous Weapons (where there is no meaningful human control over the critical functions) is also a specific form of AWS. There is a role and a place for these more specialised terms (and as a lawyer I am partial to the acronym LAWS), but in the end we need to find a solution for the problems presented by AWS, not just LAWS.
- The importance of dealing with this issue of increased autonomy in weapons release in disarmament as well as human rights *fora* become apparent – clearly the lethal, military versions should be dealt with in the CCW, but the CCW does not have a mandate over law enforcement, and unless the Human Rights Council and other such for a deals with it, an important aspect part of the problem will remain unaddressed, and is going to fall through the cracks. The Human Rights Council has a clear mandate in its founding document and thus an obligation to deal with human rights issues, wherever they occur, also in a pre-emptory way.
- From the start of the debate the importance of the right to life in this context have been emphasised. This is no doubt correct, and also explains the close engagement of my mandate with this issue. However, I have used my interventions at the meetings over the last two years to draw attention to the importance of the role of human dignity as well, and I am happy to see that the issue is squarely on the agenda, also in today's discussion.

The rights to life and to dignity are the two main rights, but following the human rights paradigm serves as a reminder that the following rights –

unless they are legitimately derogated from - also come into the picture: This includes the right against torture and cruel, inhuman or degrading treatment or punishment. (Can a robot ever administer the “human” treatment that the law requires? The use of dogs by the police for example requires proper human supervision – meaningful human control). It also includes the right to bodily security; and the right to just administrative action (can a robot ‘apply its mind’?). These are all considerations that should be taken into account in deliberations on the implications of ethical as well as legal standards (such as the Martens Clause).

- The human rights approach places a strong emphasis on the need for accountability. In fact, and very important in the context of autonomous weapons, is that a lack of proper accountability where there is a violation of a right such as the right to life in itself constitutes a further violation of that right. This highlights the importance of the concern that AWS might create an accountability vacuum – that if human control is absent we may end up in a situation where no human being may be held responsible if things go wrong with robots, because meaningful responsibility depends on meaningful control.

2. It is well-established that human rights continues to apply during armed conflict, though the exact requirements of a right such as the right to life may be determined with reference to the *lex specialis* of international humanitarian law. I would like to discuss the implications of autonomous weapons for **the right to life and the right to dignity** in the context of two questions that I think have to be asked in deciding how to deal with such weapons:

- The first question is *can they do it?* By this I mean to what extent can they be relied upon to do accurate targeting? As a technical matter, how good are or will they be at ensuring that only legitimate targets are being hit? Can they comply with the requirements of distinction and proportionality at least as well or better than humans? Much has been said about this, so let me just say that to the extent that they cannot, it becomes a right to life issue.

The people whose right to life are at stake here are those who are not legitimate targets – those who are protected by IHL, such as uninvolved civilians. As others have much more eloquently put it than I can, it is far from certain that at

least fully autonomous weapons – where there is no meaningful human control – can be relied upon to do good targeting.

But even assuming that autonomous weapons ‘can do it’ – that one day they can make targeting decisions that are at least as good as those taken by human beings – there is a second question that needs to be asked. That is the question ‘Should they do it?’ or, to put it differently, ‘Is it right?’ The main issue here is not the plight of those who are protected under IHL, such as uninvolved civilians, as was the case with the first question, but rather the implications of these weapons for legitimate targets. Is it right that humans, even if it is in other respects lawful, will be selected for targeting by machines?

This is in part a right to life question: some would say it is inherently arbitrary for machines to take life and death decisions, even where legitimate targets are concerned. This issue needs to be explored further. But for our current purposes let me just say that it is very much also a question of human dignity.

When we are dealing with fully autonomous weapons, the question arises whether the dignity of those targeted is not affected in a fundamental way, in the same way that it is a well-established rule of international law that even those who are sentenced to death in full compliance with the law may not be executed in an inhumane way, e.g. through beheading. It seems to me that such actions potentially affect the dignity of all of us, because the very value of human life is compromised. The flipside of a living a dignified life is dying a dignified death. And the manner in which one dies – or in which we can all potentially die – including how decisions are taken about life plays an important role in this regard.

A human being in the sights of a fully autonomous machine is reduced to being an object – being merely a target. This is death by algorithm; it has also been called ethics by numbers. In fact the human is reduced to the two numbers of the digital word – zeros and ones. While war is at best already an undignified business, it traditionally at least involve some kind of deliberative process, some kind of decision-making whether a particular attack is really necessary, in which human consideration may – even if often in practice not – play a role.

IHL of course sets the tone – the first question is whether it is lawful. But IHL developed in the context and in my view relies on the assumption that there will be a second-order determination as well. A human will decide whether to go ahead and actually initiate the attack. The hypothesis that machines may one

day comply better with IHL standards – doubtful as it is on its own terms – does not serve as an argument why the human can effectively be taken out of the decision-making loop. As is often said, something may be lawful but be awful, and a human is needed in the loop to prevent that from happening..

A world where the function of pulling the trigger is delegated to machines is a world without hope: hope in some measure of mercy; perhaps hope in benefitting from some measure of human error or shortcoming during the targeting. Machines cannot fathom the importance of life, and the significance of the threshold that is crossed when a life is taken.

Ron Arkin has said we should think about LAWS as a specialised kind of smart bomb, which up to a point may be true. However, the concern is that beyond a certain point – where humans are not in effective control - they may be simply *too* smart and perhaps in that sense also too predictable. It is no longer a case of humans using them as tools to achieve our purposes; the true picture is more akin to them using us for their purposes.

The above refers to the dignity of those who are targeted. However, another form of dignity is also at stake. If the decision whether to use deadly force is taken out of the hands in whose name it is being done, they cease to be moral agents: people who take a decision and assume responsibility for it. It is no wonder that so many military officers are reluctant about the introduction of AWS. Their dignity is on the line.

So far we have dealt with the use of AWS in armed conflict. There are further considerations why the use of AWS is even more problematic in law enforcement. The use of unmanned systems such as drones in law enforcement in general is already problematic, because police officers have a duty to protect the public, and in order to do so they generally need to be on the spot; they have to be hands on. But if policing becomes so depersonalised that the police officers are not only physically but also psychologically absent from decisions to use force, the very nature of policing and its basic assumptions are challenged.

Moreover, law enforcement requires a much higher level of personalised decision-making on the use of force than armed conflict: in many cases it is necessary to establish whether a person who seems to pose a threat actually intends to execute that threat. This is in essence a decision requiring human judgement.

3. So, what is the test to determine whether AWS are acceptable, in armed conflict or in law enforcement? Some AWS have a very low level of autonomy as regards the critical functions of force release; in other cases humans exercise little control. We saw above that the problems concerning the right to life and the right to dignity emerge when humans no longer exercise meaningful control.

Another way of putting it, is to ask the question whether the weapons in question are tools in the hands of humans, or the other way around. To use them as a tool is to exercise meaningful human control.

In my view **the test is that AWS may be accepted only to the extent that they are *good tools***. Both of these words carry significance.

They must be *tools* in the sense that humans use them to pursue their own objectives. Posing the requirement of meaningful human control is just another way of saying that AWS are acceptable only insofar as they are tools in the hands of humans.

But there is also a qualification in the case of those AWS that are used by humans as tools. They must be *good* tools. Where humans exercise meaningful control over machines in taking targeting decision, the expected outcome in terms of accuracy of targeting should clearly not be worse than if they were not used. But does the outcome have to be better? Or is it sufficient if the machines are as good at doing this as human beings?

I would argue that the machines have to do at least as good as humans in targeting, but the argument can be made that they should do even better than humans. Heightened technology gives increased control, and with greater control comes greater responsibility. We generally require specialists, medical and otherwise, to meet higher standards of proficiency.

So, for example, if a drone has a computer function to help its operator to select a target, such targeting should be more accurate than would otherwise be the case. Merely being as accurate as when humans act on their own means it is not a good tool and there is no reason why it should replace humans.

Cases where a level of autonomy short of full machine autonomy can help save lives in a manner which preserves human dignity can readily be imagined, and in such instances AWS can and should in my view, if available, be used. Let me give an example from a law enforcement perspective, which may resonate with the fact that this week it is one year after the 276 Nigerian girls disappeared.

Assume that during a hostage situation the lives of the hostages depend on whether snipers who have been deployed release fire simultaneously. If their rifles are connected by a computer which releases force only once all of them have a clear shot at armed hostage takers, this may allow them to get the hostages out unharmed. While the computer releases the force, this seems to me a clear case of meaningful human control being exercised, and thus a legitimate use of AWS.

It is likewise easy to imagine scenarios during armed conflict where technology is used in the targeting process, which can save lives. As long as meaningful human control is retained, this should be welcomed.

The use of technology in such a case does not undermine the dignity of those targeted, because humans take the decision, and for the same reason the dignity of those by whom it is used is also not affected. In fact their dignity can be enhanced to the extent that they are able to function better as moral agents who want to save lives – in the above case the lives of the hostages. Moreover, should something go wrong human responsibility can be assigned. Such uses of AWS as good tools should be encouraged.

4. The above implies, however, that if there is no meaningful human control, and the AWS are thus fully autonomous, they may not be used. In the last place, it may be useful to unpack the following question further, which I think is the central and most difficult question posed by AWS: **If a fully autonomous weapon can save lives on the battlefield by being more accurate in targeting than a human being, should it still be rejected?**

As was alluded to earlier, whether targeting by fully autonomous weapons will ever be superior to that of human beings given the kind of values judgements that are involved, is in dispute. However, let us assume for the moment that it can be done. In the terms used above, in such a case autonomous weapons will no longer be “tools”, but their targeting is “good”. Does the latter not cancel out the former?

Does the right to life not mandate and in fact require the use of such technology if it is available? The right to life is often described as the “supreme right”, and should surely trump softer considerations such as those about dignity outlined above?

This scenario clearly represents a dilemma. The right to life and the right to dignity appear to be in conflict. Which one should prevail? Given my role as mandate holder on the right to life, my first inclination is to say that the right to life should always prevail, and anything that can save lives should be supported.

I have given this question a great deal of thought. As you may know in my 2013 report I did not propose an outright ban, but a moratorium, to allow time for the international debate to develop, and frankly for people like yourselves and myself to make up our minds. It seems that debate has indeed developed, and has yielded some significant results. Of particular importance is the emerging consensus around the notion of meaningful human control: At least in the context of armed conflict autonomous weapons may be permitted if humans retain meaningful control, but should be banned where that is not the case.

Against the above background it seems that we have to admit to ourselves that even if such weapons can save lives, if they are fully autonomous that is a benefit we have to do without.

The idea of the indivisibility and inter-connectedness of rights means that one right cannot automatically trump all other rights. The right to life and the right to dignity, for example, are both fundamental rights, and the question which right should prevail must be determined on a case by case basis in the light of the circumstances. I would argue that where autonomous weapons are concerned, the case for the right to dignity is so strong in that at some point it prevails over the right to life. Saving the lives of some people by allowing others to be killed by machines that are not under human control involves using the latter as a means to an end and is an unacceptable affront to human dignity.

It may be more accurate not to talk about the right to life being trumped in such a case by the right to dignity, but rather to say that the right to life should be interpreted in terms of the right to dignity, because the right to life properly understood is the right to a dignified life.

It is not merely a question of protecting the right to dignity. On closer inspection, even if fully autonomous weapons can in specific cases lead to more accurate targeting, that may also compromise the right to life as traditionally understood. As was pointed out earlier, a lack of accountability for violations of the right to life is in itself also an independent violation of that right. And as we saw earlier, there are serious concerns about the extent to which there will be accountability when things go wrong with AWS. Without human control it is

difficult to imagine human responsibility. As a result, fully autonomous weapons may also threaten the right to life.

What about the use of LAWS in carefully defined spaces, for example where there are no civilians? Even assuming that the problem of those who may surrender or be wounded or otherwise *hors de combat* can be solved (which is not an inconsiderable problem), what could be wrong with allowing the targeting in such cases to be autonomous? Some would argue that meaningful human control is not in fact being exercised in such a case, by the operator selecting such a circumscribed area of deployment: that in effect the necessity and proportionality decision can be said to have been taken by a human.

This may be part of the answer, but it emphasises why it is so important to develop the concept of meaningful human control further. Clearly considerations such as how much time elapses after the machine is activated and before force is released will also play a role.

To **conclude**: it seems to me that we are getting closer to an answer to the question how to deal with AWS: As long as they are good tools, in the sense that humans exercise meaningful control over them, they can and should be used in an armed conflict situation. There is significantly less room for their use in law enforcement, where it will be difficult to outperform human beings. If they are no longer tools in the hands of humans, they should not be used.

Considerations of IHL, human rights, IHL and ethics call for fully autonomous weapons to be banned.