Meaningful Human Control, Artificial Intelligence and Autonomous Weapons

Briefing paper for delegates at the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)

Geneva, 11-15 April 2016

This paper has been drafted by Dr. Heather Roff and Richard Moyes in the context of a grant awarded to Arizona State University, in partnership with Article 36, by the Future of Life Institute (www.futureoflife.org) to further develop thinking on 'meaningful human control' as a conceptual approach to the control of artificial intelligence in the context of autonomous weapons systems.

Heather M. Roff, Ph.D.
Senior Research Fellow, Department of Politics & International Relations, University of Oxford
Research Scientist, Global Security Initiative, Arizona State University
Cybersecurity Fellow, New America Foundation

Richard Moyes,
Managing Partner, Article 36
81 Rivington Street, London, EC2A 3AY, United Kingdom
www.article36.org
info@article36.org
@Article36

Introduction

With the recent rise in concerns over 'autonomous weapons systems' (AWS), civil society, the international community and others have focused their attention on the potential benefits and problems associated with these systems. Some military planners see potential utility in autonomous systems - expecting them to perform tasks in ways and in contexts that humans cannot, or that they may help to save costs or reduce military casualties. Yet as sensors, algorithms and munitions are increasingly interlinked, questions arise about the acceptability of autonomy in certain 'critical functions,' particularly around identification, selection and application of force to targets. These concerns span ethical, legal, operational and diplomatic considerations.

The Campaign to Stop Killer Robots and other initiatives, such as the 2015 Open Letter by members of the artificial intelligence community, strongly oppose the development and deployment of certain AWS and call for a ban on uses of this technology. In response to the calls from civil society and academics, the international community and the diplomatic sphere have taken notice. For the past three years, the UN Convention on Certain Conventional Weapons (CCW) has held informal expert meetings amongst states to consider the implications of 'Lethal Autonomous Weapons Systems.' Moreover, the International Committee of the Red Cross (ICRC) hosted two of its own expert meetings on AWS. In an attempt to understand the implications of autonomous technologies, including but not limited to AWS, the UN Institute for Disarmament Research (UNIDIR) has also convened a number of expert discussions leading to various reports, and numerous other think tanks and institutions around the world have also convened workshops and meetings on the same or similar issues.

However, despite all of this engagement, the discussion of AWS is still characterized by different uses of terminology, different assessments of where the 'problem' issues really sit, and divergent views on whether, or how, a formalized policy or legal approach should be undertaken.

Nevertheless, amidst the developing discussion, the concept of 'meaningful human control' (MHC) emerged as one point of coalescence. Primarily, it has been used to describe a threshold of human control that is considered necessary; however, the particulars of the concept have been left open so as to foster conversation and agreement. It is necessary, however, to address in more detail the content of this principle. This paper seeks to do so by offering a framework for meaningful control to a multi-stakeholder audience from a diverse set of professional and academic backgrounds.

The development of 'meaningful human control' as a policy approach

At its most basic level, the requirement for MHC develops from two premises:

1. That a machine applying force and operating without any human control whatsoever is broadly considered unacceptable.[1]
2. That a human simply pressing a 'fire' button in response to indications from a computer, without cognitive clarity or awareness, is not sufficient to be considered 'human control' in a substantive sense.

From both of these premises, questions relating to what is required for human control to be 'meaningful' are open, as well as how far away in distance and/or time a human has to be from an act in question for there to be 'human control.' Given the openness of these questions, MHC represents a space for discussion and negotiation. In this paper we will often use the term 'human control' as synonymous with 'meaningful human control' – as the word 'meaningful' is considered to function primarily as an indicator that the concept requires further colelctive definition in policy discourse. Approaching the challenges of autonomous weapons from this entry point situates the discussion from a positive point, one where many states and civil society agree – in broad terms if not necessarily in detail.

As noted, the concept presents substantial space for divergent opinions on where the boundaries of necessary human control might lie. Possible issues for further negotiation include:

x Meaningful human control over what? Is the concept being applied to the technology itself or to the wider situation within which a technology might be applied? Article 36 has called for meaningful human control "over individual attacks", but other actors have used the term differently.
x To what extent can aspects of 'human control' be programmed into autonomous technologies?
x To what extent should current practice regarding human control shape normative expectations for the future - recognising that there are already limitations to levels of human control exerted in existing military systems during the application of force, as well as to the application of force itself?
x Is there a fundamental threshold necessary for human control and can we assess technologies to determine whether they fall on one side of that line or another?

Ultimately we also recognize that many answers to these questions and others are likely to be 'political,' that is, different actors might prefer different formulations or categorizations based on wider considerations or interests. However, despite this fact, we hope to provide positive content for the international community and civil society as human control becomes a more central focus for discussions.

Conceptualizing human control in socio-technical systems

One starting point for approaching MHC is to consider the principles by which 'human control' over technological processes, and the systems within which they are embedded, might be understood in general terms. The elements suggested here are not proposed as being definitive or exhaustive, but rather to provide working tools that can inform discussion towards policy development. The presentation of such elements here seeks to build up common understanding of the necessary human control required for the operation of weapons systems, recognizing that the decision about where to draw the line over the permissibility of weapons systems that incorporate autonomy will be political, rather than purely technical.

In general terms, human control over technology is enhanced if:

The technology is predictable
x The technology developed should be predictable in its functioning, within certain understood parameters.

x This is linked to issues of design, production, storage and maintenance, and to the provision of accurate information into the wider system.

The technology is reliable
x The technology developed should be designed for reliability.
x The technology developed should be designed for graceful degradation in the case of malfunction. Systems designed to degrade to prevent catastrophic failure constitute one such way that humans might design for control when exogenous events or shocks occur.

Predictability and reliability often are paired when discussing the functioning of a system or artifact. This is so because predictability and reliability are the primary metrics by which humans can measure whether their creations are continuing to function as intended. However, given the certainty that all tools break, one must be sensitive to designing for safety once intentional or unintentional malfunctions occur. In addition, it is important to recognize that 'predictability of outcomes' is also dependent upon understanding or controlling the spceific context within which a technology will function.

The technology is transparent
x The technology ought to be designed so that if necessary, one can interrogate the system to inform the user or operator about the decisions, goals, subgoals or reasoning that the system used in performing its actions.
x There should be clear goals, subgoals and constraints emplaced on each system, and it must be possible for human operators to understand these.
x Clear and intuitive design of systems and user interfaces should encourage responsible and intended use; designed for the practical user and not for an ideal user in a lab.

Transparency of a system is one way that designers and users can interrogate it to ensure that the system comports and upholds the goals, subgoals and constraints emplaced by the designers, planners and operators. There ought to be opportunities for feedback between a system and its operator so as to ensure that the human operator has sufficient degrees of situational awareness with regard to the system's operation in its environment, as well as whether the system is functioning within appropriate parameters. If a user's goal is modified, due to changes in the context or of that user's intent, then human control would require that the system be aligned with that new goal, with transparent functions providing assurance that this alignment occurred.

The user has accurate information
x The user(s) should understand the outcomes that are sought (i.e. to what purpose the technology is being used).
x The user(s) should understand the technology and the process that will be applied.
x The user(s) should understand the context within which that technology will function.

If we conceptualize technology as a tool for translating user intent into outcomes in a particular context, information on these three elements – the intent, the technology, and the context – become of critical importance to an assessment of whether that technology is under effective human control.

Predictability, reliability and transparency of technology all contribute to a user's capacity to understand the technology that they are working with. Yet these technological characteristics cannot by themselves ensure necessary human control, which is dependent upon a wider understanding of the outcomes that are being sought (including outcomes that are to be avoided) and of the context within which the technology will operate (i.e. those things in the wider environment that it may interact with).

Information on all of these elements is likely to be produced by and in a wider political or socio-technical system. In turn, confidence in any information is likely to be driven by numerous factors, including proximity to the source, past reliability of the source, attitudes towards the system(s) itself, verifiability of information, and/or transparency of system functions. Thus a sufficient level of confidence in information to ensure human control may itself be managed by other wider or interacting systems. These other systems, however, may also provide opportunities for individual human judgment, but they may also produce challenges for human judgment, such as where over-confidence in systems can produce forms of bias.

There is timely human action and a capacity for timely intervention
- x A human user is required to initiate the use of a particular technology while the contextual information they are acting upon is still relevant.
- x Although some systems may be designed to operate at levels faster than human capacity, there should be some feature for timely intervention by either another system, process, or human.
- x 'Timely' may range from picoseconds to hours or days, depending on the technology and the context (including the level of inadvertent harm that might be caused), and structures of accountability within which the technology is being used.
- x Accountability is conceptually and practically linked to the potential for timely human action and intervention in that accountability resides with some human or set of humans.

Action by a person (or persons) seems to be necessary for human control. At the most basic level, such action might involve the starting and stopping of processes on the basis of contextual reasoning and judgment. As framed here, the most significant actions are those that tie the information or data being acted upon to the technological process being applied. While we may more easily conceptualize this as a single person's judgment and action, in reality there are likely to be different people undertaking a variety of actions at different points in a process (for example, in the maintenance of the technology to ensure reliability, or the production of information about the context in which it will be used). The key is to ensure that a person or persons are capable of action and intervention, and if inaction or nonintervention occurs there these individuals are identifiable for accountability measures.

There is accountability to a certain standard
- x Accountability should reaffirm that a human person or persons are responsible for processes initiated.
- x Accountability should condition the socio-technical system by ensuring that people understand that they will face consequences for their actions or inactions.
- x While primary accountability may lie with the person(s) whose actions most directly tie together the system, accountability must

also come to bear upon wider systems or organizations that produce such socio-technical systems and artifacts.

Accountability is an ex post process to locate responsibility or liability with human agents, but it also establishes a framework of expectation that can guide human agents to align their behavior with expected and appropriate standards. Standards for accountability, moreover, need to ensure that responsibility and liability will be apportioned equitably, and that sanctions will be applied that are commensurate with the wrongdoing (whether intentional or inadvertent) and with the severity of harm that may have been caused.

On the basis of the analysis above, the key elements for human control are:

- x Predictable, reliable and transparent technology.
- x Accurate information for the user on the outcome sought, operation and function of technology, and the context of use.
- x Timely human action and a potential for timely intervention.
- x Accountability to a certain standard

Meaningful human control in the context of the use of force

We can also consider and bring to bear the concept of human control outlined above to the context of the use of force. We can do so within a broad, chronological framework where different elements have greater prominence at other points in time.

Whilst it is in the use of any autonomous weapons system (AWS) that the key concerns regarding a lack of human control in the critical functions reside, the lack of human control at this layer cannot be fully addressed without consideration of the processes of design and development that precede it (and that it draws upon), and the 'ex post' structures of accountability that encompass it.

Although approached in a broadly chronological order here – from weapons development, through use, to systems of accountability – each of these layers also serves to shape and to condition the others, and consideration in all of these areas should inform next steps for ensuring meaningful human control in the future. We thus present a three-layered approach to MHC: ante bellum; in bello; and post bellum. At each layer, there are systems, processes, and doctrines designed to uphold human control, and so it is appropriate that we view it beyond limited engagements.
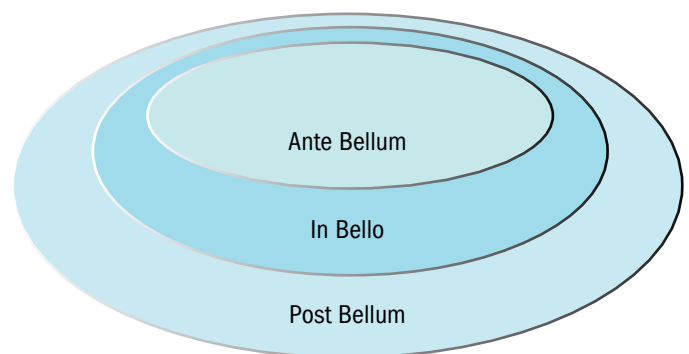

Fig. 1. Human control needs to be embedded through mechanisms operating before, during and after use of technologies in conflict.

Design, development, acquisition and training

The first layer for consideration encompasses all of the processes for control 'ante bellum' or before the start of hostilities. Of particular concern are design, development, acquisition, and training in relation to the use of force. Applying force implies that a main consideration is how human control is conceptualized, embedded or indoctrinated during these processes, especially in relation to weapons systems. In so far as weapon systems are designed, marketed and acquired as tools for achieving certain human ends, and recognizing that weapons systems may also produce unwanted outcomes, it is to be expected that human control is brought to bear during the processes of design and development.

'Responsible innovation' is an approach in science and technology studies that recognizes that 'science and technology are not only technically but also socially and politically construed.'[2] As a result, it recognizes that science, technology and innovation are situated in socio-political worlds, and that scientists and researchers have particular responsibilities to ensure that their work does not have harmful consequences. Recognition of the role of scientists and researchers in shaping both the technological and socio-political trajectory of weapons' development has a significant history. For instance, in 1955 Bertrand Russell and Albert Einstein issued a Manifesto calling for states to renounce nuclear weapons (specifically the hydrogen bomb) and to find peaceful resolutions to conflict. Their call came on the heels of extensive engagement by scientists and physicists in the development and production of these weapons. While the resulting Pugwash conference did not end research or development of nuclear weapons, later attempts at responsible research and innovation did succeed. In the 1970s research on recombinant DNA was halted due to fears about what the findings might unleash. Biologists saw the dangers coming, and in 1975 issued a moratorium on its development. More recently still, the US' Defense Advanced Research Project Agency (DARPA) called together an ethics advisory board to oversee research pertaining to plant-DNA manipulation. These examples simply illustrate that to different degrees moral or ethical imperatives can be brought to bear in the processes of technological development.

Given that autonomous weapons systems (AWS) are an emerging technology, there are many uncertainties about the risks involved with their future development. Also problematic is that while many of the constituent parts of AWS are dual use, much of what we know about aspects of current AWS capability is either classified or speculative. Thus public dialogue that might facilitate "value sensitive design" is potentially curtailed due to national security or commercial confidentiality concerns.[3] So if the primary concern regarding such technologies (whether in their use or by design) is framed as a lack of meaningful human control, then it is fair to say that there is not yet a common agreement on the form of human control that designers and developers should be working to embed. There is even less agreement on any doctrinal changes or challenges that AWS may pose to existing military command structures.

In the general framework laid out in the previous section, issues of 'predictability', 'reliability' and 'transparency' were all raised as allowing for more or less substantive human control. Yet it was also noted that 'predictability' is not just a characteristic of a technology, but also of its operation in specific, understood circumstances (this will

be discussed further in the section below on 'human control during attacks'.) The implication of this is that any technology must also be designed for and accompanied by guidance and constraints on the contexts in which it can be used if operational predictability is to be found during operational functioning.

How then might computer scientists, roboticists, electrical engineers and weapons manufacturers, and acquisition specialists incorporate responsible innovation and design into their work? Particularly, how might they design for human control when we know that certain aspects of the systems developed may operate outside of human physical control?

In such a context, it is pressing to build agreement across the social and political context within which design and development processes are taking place that the requirement for meaningful human control over the application of force must be embedded into those design and development processes. How that requirement might be embedded could develop through further ongoing dialogue, but unless an external expectation for meaningful human control can be brought to bear on such processes it may be unreasonable to expect "responsible innovation" to happen on its own.

Human control during attacks

A second layer for consideration relates to human control during the conduct of hostilities ('in bello'). In particular, MHC is concerned with maintaining human control at the level of 'attacks.' 'Attack' here is used as a term in the context of international humanitarian law - the legally binding framework that regulates the conduct of hostilities - and can be thought of as distinct from operational or strategic planning within the military. In many respects it is the requirement to ensure meaningful human control within this layer, over attacks and specifically over the critical functions of identification, selection and the application of force to targets, that drives the need to embed human control within the wider frameworks of systems of development and accountability.

Human control here looks to how human commanders take the products of the first layer, in terms of technology and guidance, and apply them to specific contexts of hostile actions in time and space. How and why such contexts are chosen, of course, are themselves the product of wider systems of information gathering and decision-making. However, 'in bello' human control is concerned with a human commander weighing her expectations of using a certain technology in a specific context against the risks of unwanted outcomes (while recognizing that there are thresholds for accepting certain risks). These human evaluations and judgments are necessary for adherence to the law. The commander's understanding of the military objective being sought, of the particular weapons system deployed to bring about the achievement of that objective, and of the context in which the weapon will be used (situational awareness and intelligence), all have bearing on the commander's judgment of predictability. In general, a less predictable, reliable and transparent weapon technology, operating in a more complex environment, over a wider area and for a longer period of time will likely reduce a human commander's ability to meaningfully predict outcomes.

There are three levels generally associated with military action: the tactical, the operational and the strategic. Meaningful human control

over 'attacks' should be understood to require human control at the lowest level at which human legal judgment must be applied and as such resides at the tactical level of warfighting.

Another way to think about this is that MHC precludes a commander at either the operational or strategic levels from meeting the preconditions for 'control' over attacks. As Article 57 of Additional Protocol I of the Geneva Conventions imposes a positive obligation on parties to a conflict, particularly 'those who plan or decide upon an attack' to take constant care and precautions with regard to the civilian population, civilians and civilian objects, we can infer that human commanders have a duty to maintain 'control' over attacks by engaging in precautionary measures. Furthermore, we can also infer that the drafters' intent at the time was to require humans (those who plan or decide) to utilize their judgment and volition in taking precautionary measures on an attack-by-attack basis. Humans are the agents that a party to a conflict relies upon to engage in hostilities, and are the addressees of the law as written.
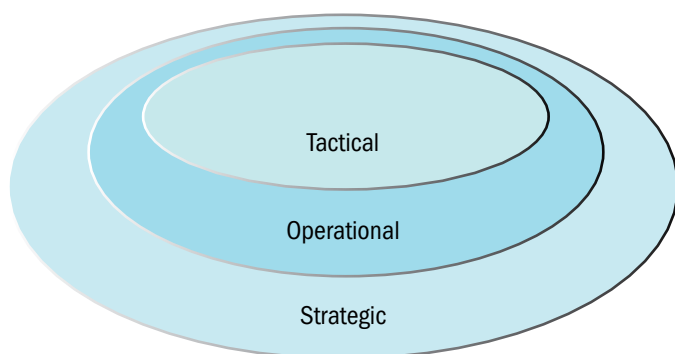

Fig. 2. Meaningful human control needs to be applied over attacks at the tactical level of warfighting, as well at other levels.

As the obligation resides at the level of attack, which implies a tactical unit of analysis, this means that one cannot satisfy the MHC criteria by pointing to a human at either the operational or strategic level and claim that orders given here are sufficient to show control over the tactical levels. As such, MHC precludes AWS from moving from one 'attack' to another without so being ordered by a human to do so, and without each individual attack being subject to human legal judgments. The term 'attack' in these legal rules should be understood as providing a unit of analysis, where human judgment and control must always be applied. To abandon such an understanding would be to undermine the structure of the law as a framework addressed to human legal agents – whether individually or collectively.

However, the parameters of what constitutes an individual "attack" are not defined in practical terms. It is generally recognized that an attack can involve multiple applications of force to multiple target objects. The extent to which these objects should be geographically proximate to each other, and the duration over which a use of force may constitute an individual attack, are all open questions to some extent. Whilst existing practice may suggest some boundaries to what might be considered an "attack," AWS may also challenge these notions, as they may offer the potential to strike at geographically disparate objects that fall within a certain target classification. Such dispersion would present further challenges to a commander's access to information on the specific context within which force will be applied.

As was indicated in our general analysis of control, contextual understanding – information on the geographic space and the time within which a technology will be used - is vital for effective control over that technology. Broadening the concept of an 'attack' risks diluting the information available to human commanders as a basis for their legal and operational judgments to the point where their ability to predict outcomes becomes either nonexistent or minimal. Thus, it is not enough to say that the use of such weapons 'needs to comply with the law because the boundaries of such legal terms as 'attack' are not fixed and are open to different interpretations and reinterpretations by the use of new technology over time.

If the technological capacity afforded by 'autonomy' pushes 'attack' towards being conceptualized more and more broadly (e.g. pushes out from the tactical to the operational and strategic layers), then a requirement for meaningful human control should be used to counteract this move so as to avoid a progressive dilution of the law and its requirements for human judgment and application. It is for this reason that articulating a requirement for meaningful human control is an essential initial building block for policy and legal responses to prevent the development of an acceptable autonomy in the critical functions of weapons systems.

Command structures and accountability

The third and final layer of analysis relates to accountability. As noted previously, meaningful human control links the need for responsible design and use to systems of accountability. They are conceptually and practically linked. In this way, we see how the other two layers flow into and require the third. For when ante bellum or in bello mechanisms fail, there is a need for accountability. Regardless if one views accountability measures as necessary for punishment, deterrence, social utility or as a means to return to the status quo ante, such measures are necessary features of a system of human control and law.

Questions pertaining to how traditional notions of accountability would be challenged by the deployment of autonomous weapons remain of importance in this debate. Some argue that commanders may not be held responsible for the war crimes committed by an AWS, and that the most a commander might be responsible for would be recklessness (although international law does not have a legal framework to prosecute individuals for recklessness during hostilities).[4] Others side-step this problem and claim merely that "the goal must ultimately be to ensure the autonomous weapon functions in a manner that, like the soldier, is subordinated to the will and parameters imposed by responsible command" without establishing how to ensure that such a system – which is not a moral or legal agent – can be subordinated in the same way that a human warfighter is subordinated.[5] Roff claims instead that for the near term advanced autonomous weapons might be considered analogous to marine mammals used for intelligence, surveillance and reconnaissance, diver detection, and mine location.[6] On this reading, operators or commanders may be held accountable for a dereliction of duty by a failure to appropriately train and care for one's animal. Others posit that if advanced systems were allowed in the future it may be that no person could be held accountable in a way that is adequate to the possible outcomes, and so there would exist an "accountability gap."[7]

Much depends, however, on whether or not one accepts an obligation to assert meaningful human control over direct attacks, and thus over the use of AWS. With an established acceptance of a requirement for meaningful human control, issues of accountability become more straightforward because the person(s) exerting control are established as responsible through an accountability system within which they are operating. Many questions regarding accountability stem from a blurring of the approach to AWS, an approach that moves from treating them as tools to treating them as moral or legal agents. This movement itself tends to breach the requirement for case-by-case human legal judgment and control over attacks that we see as implied by existing law.

'Control' is already recognized as a vital element within existing accountability structures. For instance, in cases of a failure of command responsibility, there is the requirement of a crime having been committed. A commander's responsibility, in the case of negligence, is where she failed to prevent or to punish those under her command, and command requires the 'effective control' of a commander over subordinates. If a commander does not have effective control, then she cannot be held responsible under a doctrine of command responsibility for negligence. Failure to prevent or to punish those outside of one's control would rather be a doctrine of strict liability. If a commander directly orders the commission of a crime, then she is held responsible for her direct order, as well as for the crimes of her subordinates. If we view AWS as tools, and not as agents, then we have the opportunity to use this existing framework to operationalize requirements for human control over direct attacks. If this is insufficient, then there is opportunity here to refine the responsibilities of commanders and operators in line with existing legal notions like strict liability, recklessness or dereliction of duty.

Drawing boundaries: policy and technology

The process of policy development generally requires the adoption of certain boundaries or thresholds and their related categories so that objects or behaviors can be managed. Creating such boundaries requires drawing lines that serve to simplify a complex reality. The implications of the sections above are that broad 'key elements' of meaningful human control can be delineated, and that establishing a requirement for meaningful human control in the context of autonomous weapons systems is a necessary first step towards ensuring those key elements are maintained as military technologies (and the structures within which they are embedded) develop in the future.

While diplomatic responses to the concept of meaningful human control tend to fixate on the term 'meaningful', this is generally a failure to recognize that that specific word merely indicates a need for the policy community to undertake the work of delineating what form of human control is necessary. This process could draw, in normative terms, upon the general principles of control suggested here as cumulatively constituting meaningful human control, before during and after the use of force – sufficient predictability, reliability, and transparency in the technology, sufficient confidence in the information that is guiding the human judgments being made, sufficient clarity of human action and potential for timely intervention and a sufficient framework of accountability.

Specifying the level of 'sufficiency' in all of these areas may be difficult in detailed terms. Nevertheless, categories of technology may still be assessed against these considerations. Technological boundaries, such those between 'automation' and 'autonomy' might be boundaries that in turn represent different capacities for predictability, for reliable information on the context of use, for timely intervention or for the coherent application of accountability. Similarly, any broadening of the legal concept of 'attack' could also be challenged against the tests that these normative requirements present. The latter is particularly important because it should remind us that without asserting a requirement for human control, to some standard, the legal framework itself is a malleable framework. Its malleability is both a benefit and potential harm, for as a benefit it can change when the need arises, but it can also change through (un)intentional abuse. Simply asserting a need for legal compliance is not enough when key terms might be interpreted so openly in the context of AWS as to render hollow any claims that human legal judgment is being applied.

Consideration of the key elements required for meaningful human control should provide a starting point for any assessment of developing technologies in the context of autonomous weapons systems. The positioning of definitional boundaries and determinations of what form or extent of human control is considered sufficient or necessary will represent political choices, with different actors favoring different options based on different assessments of their wider interests. However, developing the basic framework against which such assessments might be made is essential to such subsequent processes of analysis.

END NOTES

1 "Machine" here encompasses software systems, artificial agents, and robotics, as well as combinations thereof.

2. Stigloe, Jack, Owen, Richard and Macnaghten, Phil. 2013. "Developing a Framework for Responsible Innovation" Research Policy, 42(9): 1569.

3. van den Hoven, J., 2007, in IFIP International Federation for Information Processing, Volume 233. The Information Society: Innovations, Legitimacy, Lthics and Democracy, eds. P. Goujon, Lavelle, S., Duquenoy, P., Kimppa, K., Laurent, V., (Boston: Springer), pp. 67-72.

4. Ohlin, Jens. 2016. "The Combatant's Stance" International Law Studies, 92: 1-30

5. Corn, Geoffrey S. 2014. "Autonomous Weapon Systems: Managing the Inevitability of 'Taking the Man out of the Loop' "Available at SSRN: http://ssrn.com/abstract=2450640 or http://dx.doi.org/10.2139/ssrn.2450640

6. Roff, Heather. "Command Responsibility and Lethal Autonomous Weapons" Unpublished and onfile with the author.

7. Human Rights Watch. 2015. "Mind the Gap: The Lack of Accountability for Killer Robots" https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots