



CCW REPORT

Civil society perspectives on the CCW Group of Governmental
Experts on lethal autonomous weapon systems
13–17 November 2017



Reaching Critical Will

15 November 2017
Vol. 5, No. 3

IN THIS ISSUE

1 | Losing control: the challenge of autonomous weapons for laws, ethics, and humanity

The CCW Report is produced by the Reaching Critical Will programme of the Women's International League for Peace and Freedom (WILPF). WILPF is a steering group member of the Campaign to Stop Killer Robots.

The views expressed in this publication are not necessarily those of WILPF or the Campaign.

Editor: Ray Acheson

www.reachingcriticalwill.org
info@reachingcriticalwill.org



LOSING CONTROL: THE CHALLENGE OF AUTONOMOUS WEAPONS FOR LAWS, ETHICS, AND HUMANITY

Ray Acheson | *Reaching Critical Will of Women's International League for Peace and Freedom*

During Tuesday's session of the group of governmental experts (GGE) on lethal autonomous weapon systems (LAWS), participants engaged in discussions on military effects and the legal and ethical dimensions of autonomous weapons. Whilst the experts speaking on the military effects panel were generally more enthusiastic about the potential benefits and utility of LAWS than their counterparts on the legal/ethics panel, or the technology panel the day before, common themes did emerge amongst all these discussions.

One theme centred on the use of terminology, particularly around the terms autonomy or autonomous, with a variety of suggestions about what is the most appropriate way to delineate the systems under consideration at this meeting. This inevitably leads to a concern about definitions and about the scope of weapons under consideration, with some participants advocating for a narrow scope dealing with only future weapons and others for a much broader scope that incorporates existing systems that operate with some degree of autonomy.

Apart from these definitional or terminological issues, perhaps the most crucial issue under debate is the degree to which autonomous systems operating outside of "meaningful human control" can comply with international humanitarian law (IHL), and, related to this, what ethical principles could possibly be programmed into an autonomous system that would help facilitate and interpret such compliance. The idea that a machine could be programmed with ethics is anathema to some participating in this debate—to whom the notion of what it is to be a human being is considered incompatible with the concept of a machine—whilst others seem to believe that such programming might be possible.

At the end of the day, what is clear from a legal, ethical, and operational standpoint is that there are dangers and risks inherent in the concept of a truly auton-

omous weapon—one that would operate outside of human control. While a minority of participants in these discussions have asserted some operational benefits of such systems, the vast majority agree that without human intervention, robotic weapons will struggle or likely fail to comply with existing laws and with human ethics, which creates significant challenges for legal accountability, protecting civilians and civilian objects against violence, and preserving peace and stability.

The following is a summary of the discussions that took place under both panels; it is not a comprehensive record of all questions and perspectives. Unfortunately, women made up a vast minority of those experts speaking on the panels so far: only one woman on each of the technology and military effects panel, and two on the legal/ethical dimensions panel.

Panel 2: Military effects

Brigadier (Armament corps) Patrick Bezombes, France, divided automated systems into four levels: teleoperated, supervised, semiautonomous, and autonomous. Each moves further away from having human command and control. Currently, the first three levels are already being developed and deployed by many countries. Bezombes recommended the GGE focus only on future technologies such as fully autonomous weapons; the main characteristic of which is the lack of subordination to military hierarchy, lack of control, and lack of disactivation.

Professor Heigo Sato, Takushoku University, Tokyo, described four interdependent points of political-military aspects of LAWS: the diverse nature of technology application to weapon systems; the differences between objectives and military doctrines in each country; the diversity of military decisions; and dual-use control and proliferation management. He argued that whether states decide to adopt weapon review processes, a non-proliferation regime, or a prohibition on LAWS, all will

continued on next page



Losing control, continued

require various mechanisms to address the complexity of the development of these systems.

Lieutenant Colonel Alan Brown, Ministry of Defence, United Kingdom, argued that while the nature of warfare will remain violent, visceral, and prone to chaos, weapons and the world will evolve. He commented that the idea that warfare will one day fall entirely to machines is morally wrong and technologically unlikely; he noted that when it comes to using a robot in urban conflict, right now it would likely struggle to open a door let alone distinguish between a child and a combatant. He also argued that the “moral horror” expressed over delegating agency and choice to machines leading to an accountability gap is not accurate; he asserted that the only way to create that gap is by changing laws. If equipment is improperly used or bad choices are made, we hold humans accountable in the current context and this should not change. The responsibility would lie with the commander. He concluded that weapons review process is the best way to conduct assessments of LAWS.

Dr. David Shim, KAIST Institute for Robotics, Republic of Korea, said that right now AI systems are too stupid to make weapons of, which may be dangerous, but that artificial intelligence (AI) will likely advance. He argued that the key benefit of a human operator is that we can ask why certain decisions were made, and suggested programming a similar process into LAWS, where through rigorous testing we can determine why certain decisions are made. He argued that designers, operators, and commanders would need to be held responsibility for the operational conduct of LAWS.

Lieutenant Colonel Christopher Korpela, Military Academy-West Point, United States of America, argued that there are many operational advantages of unmanned and remotely operated systems, which autonomous systems would share. He also argued that autonomous weapon systems offer potential humanitarian benefits, such as less “collateral damage”. He agreed with Brown that weapons employment is the commander’s role and responsibility.

Ms. Lydia Kostopoulos, National Defense University, Washington, D.C., discussed a three-part approach to the drivers of autonomous weapons: trust of the technology, cultural acceptance, and availability. She said that to build trust in a technology, it must perform exactly as expected. She also argued that as “digital natives” (youth that are growing up with AI and other relevant technology) become commanders, their understanding of what is acceptable in the battlefield may be different from ours. Finally, she noted that a state or non-state actor can only use technology that is available, and as AI and other

LAWS-related technology becomes democratised, we will see more use of it.

The discussion after the panel presentations ranged from questions about the scope of the technologies under consideration to the tactical, strategic, and operational challenges that LAWS may pose.

Brazil, Ireland, and Switzerland all expressed concern about efforts to limit the conversation only to **future technologies**. They all noted the benefits of considering all emerging technologies with some degree of autonomy, with Switzerland noting that such technologies will help determine how systems can comply with international humanitarian law (IHL). Switzerland also noted that a narrower definition may be necessary for regulation of weapons, but not for discussion at this stage.

Brazil and Ireland were also concerned about restricting discussions to “**lethal**” **weapon systems**, noting that this is highly unusual in the context of international humanitarian law and the CCW itself. IHL incorporates destruction of civil structures, as Brazil noted, and LAWS may be effective in such material destruction as well as lethality.

Brazil, reiterating that in its views national weapons reviews are not sufficient to address the challenges of emerging technology including LAWS, argued that **developing common understandings** of this technology is necessary. Brazil expressed interest, in the context of a UN convention, to agree on a multilateral common denominator that goes beyond knowledge of specific domestic systems in order to set an international standard of what is acceptable from an IHL and international human rights law (IHRL) perspective.

Egypt, Sierra Leone, and a representative from the University of Birmingham questioned Korpela’s assertion that autonomous weapons would result in less “**collateral damage**” (aka humanitarian harm, civilian casualties). Egypt asked if bias in programming could actually lead to more civilian casualties, while the Birmingham representative noted the challenges we already see with armed drones causing civilian casualties. Sierra Leone asked that if machines cannot have common sense, as we heard in the technology panel on Monday, and even the most trusted or sophisticated systems are subject to malfunctioning, is there any reason there would not be collateral damage and who should be held responsible.

Korpela responded by reiterating his belief that the “preponderance of intelligence and information” to provide a clear picture of the battlefield helps reduce risks of “collateral damage”. He argued that the main problem with civilian casualties lies with a lack of situational awareness, and asserted

continued on next page

Losing control, continued

that LAWS could offer a more proportional and discriminate response in an urban environment, through direct engagement of a single combatant, faster decision making processes, or use of non-lethal force. He did not respond to the problem of a malfunctioning system in this context.

Japan asked if there would be any circumstances in which a commander would be **reluctant to deploy LAWS**. Brown responded that this depends on what we trust the systems to do. From a soldier's perspective, he argued, they would never want to use equipment if it is unpredictable—not just because of a moral imperative but because there is military tactical value in using a system you know is reliable. In this context, Brown re-emphasised his preference for weapons reviews, which he argued would preclude a state from producing and using weapons that are not proportional, precise, and discriminate.

In this vein, Sierra Leone noted that in previous CCW meetings on LAWS, we have been told that this technology is far away and that no military commander would delegate command and control to such systems. Now we are hearing about cost effectiveness and military advantages of these systems, so are there then situations in which a commander would use these weapons and delegate such authority?

Egypt asked about the **strategic implications** of LAWS, including its impact on parity amongst states more broadly than a tactical advantage on a certain battlefield. Kostopolous argued it is too early to say. The Chair deflected the question to other discussions yet to come.

Russia asked to what extent could a CCW protocol **control development of AI** in private companies or regulate a "grey market" in relevant technologies. Bezombes said in order to regulate the market, a technical normative reference is needed to allow all industry actors and administrations to guarantee a certain level of quality in production of AI algorithms.

Sweden asked how states could develop **methods and tests for reaching trust** and understanding for complex systems? Shim called for a new kind of testing paradigm, such as rigorous testing for all kinds of imaginable scenarios, rather than a sampling of capabilities.

China asked about the **differences between automation and AI**. Bezombes reiterated his belief that the word autonomy can be misleading, and that AI is a fashionable term that is widely misunderstood. Shim categorised automation, autonomy, intelligence separately, explaining that automation is certain kind of process or procedure through which we know exactly what will happen; autonomy is a set of rules that a machine will follow, but we don't know exactly what will happen because there is complexity and thus it is hard to predict

the end results; and intelligence is a system can write it's own rules, so human-given rules are not always obeyed and rules can be newly invented by the machine.

Uganda and Pakistan expressed concerns about **proliferation** and a potential **arms race** in LAWS. Cuba, Brazil, and Sierra Leone expressed concern about the division highlighted particularly in Kostopolous' presentation between **law abiding and non-law abiding actors**. They argued that all UN member states must be treated equally. The Chair suggested that this distinction is between states and non-state actors, rather than amongst states.

Panel 3: Legal/ethical dimensions

Ms. Kathleen Lawand, International Committee of the Red Cross (ICRC), noted that every state is responsible for ensuring that any new weapon is capable of being used and is used in accordance with IHL, which is why national weapon reviews are crucial. However, they are not necessarily sufficient to deal with LAWS, as the unique and novel characteristics of this technology may give rise to difficulties in interpretation and application of existing IHL, leading to inconsistencies between states. She also outlined the ICRC's proposal for a broad working definition of autonomous weapon systems, which is that of a weapon that has autonomy (no human intervention) in critical functions such as targeting, and that the weapon system, after being launched by a human, takes on functions that would otherwise be controlled by humans. She clarified that the purpose of this definition isn't to identify autonomous weapon systems that would be subject to concern or regulation, but to provide a baseline for discussion.

She agreed with states from the morning session that raised concerns about using the term "lethality," arguing that lethality is not inherent to a weapon but depends on how it is used. Instead, focusing on autonomy in critical functions is the most relevant approach for determining compliance with IHL. Compliance is based on degree of human involvement, she explained, not on the technology itself. In this regard, some form of human control is necessary to ensure compliance with IHL, in the ICRC's perspective. Law is addressed to humans and the legal obligations under IHL rest with those who plan, decide upon, and carry out attacks. Accountability cannot be transferred to machine, computer programme, or weapon system. Compliance with IHL may require constant supervision, as well as communication links with a human operator to permit adjustments or cancellation of a mission. Compliance with IHL, Lawand emphasized, requires a direct link between the decision or intention of a human operator and

continued on next page



Losing control, continued

the outcome of an attack. AI machine learning raises concerns due to the inherent unpredictability; thus there may be a need to develop standards in predictability and reliability in autonomy.

Ms. Marie-Helen Parizeau, World Commission on the Ethics of Scientific Knowledge and Technology, UNESCO, called for a distinction between robots in light of their autonomy, or to what degree they can make “decisions” without human control. She emphasized that the moral responsibility of taking human life cannot be delegated to a machine, because this would violate human dignity and devalue all human life. Thus for ethical, legal, and operational reasons, human control has to be kept over weapon systems and use of force.

Professor Xavier Oberson, University of Geneva, spoke about the liability of an autonomous weapon system, which by its nature would not have a “legal personality” that could be liable for crimes or misconduct. Those prosecuting such crimes would have to define a “chain of liability” from production to use of the weapon.

Mr. Lucas Bento, Attorney, Quinn Emanuel Urquhart & Sullivan, and President of the Brazilian American Lawyers Association, urged participants in these discussions to consider shifting their terminology from “autonomy” and “lethality” towards “intelligence” and “violence,” arguing this would provide a more precise framework for a working definition. On the topic of liability, he suggested that a precautionary approach to save lives is necessary, and argued that a system should be able to inform its human operator whether it is functioning properly or experiencing errors, and to request control transition from system to operator. In short, there must be communication with a human, and possibility of human intervention. In a reference to the morning’s conversation about learning to “trust” machines, he suggested that part of this is needing the machines to explain back to us what it is doing and why.

Professor Bakhtiyar Tuzmukhamedov, Diplomatic Academy, Russian Federation, asked a series of questions that arise for him in the context of LAWS, such as: who becomes a legitimate target for retaliation in the use of autonomous weapons; how can one disengage a system from a mission after it has been completed or once the weapon has been activated; how can we prevent unauthorised access; etc. He also argued that under the CCW, protocols prohibit or restrict specific weapons. Without a definition, he purported, we are not ready for a legally binding instrument. Otherwise, there is the risk of having to apply a document that reflects an agreement to disagree. Instead, he suggested, it might be worth considering non-legally binding arrangements such

as guidelines or guiding principles, that contain initial common understandings on modes of use; these could be further developed as knowledge about prospective systems is gained and enriched. He added that it might also be worth considering non-proliferation and export control approaches.

Professor Dominique Lambert, l’Université de Namur, outlined four characteristics that could be considered unique to human beings: responsibility, ability to relate to others, creativity, and compassion. A robot could imitate these principles, he argued, but without understanding what is at stake. In a quest to preserve a “principle of anthropological self-consistency,” he suggested that we must prevent the development of unpredictable systems, or systems that mean a human being would lose, dilute, or conceal the responsibility that human has for their actions.

In this regard, he suggested, at minimum we can preventively point to LAWS we don’t want on grounds of human consistency. This could imbue meaning into the phrase “meaningful human control”. It may allow for autonomous weapon systems that could help humans survive in a conflict context, provided its understood these systems are in safe, fixed space and only undertaking actions that responsible authorities have stipulated. We don’t want autonomous robots that are totally innovative that deprive us of compassion and responsibility.

In the discussion that followed, states raised a number of important questions about transparency in national-level weapon reviews, the sufficiency of existing international law,

Estonia asked the ICRC if there are other **critical functions beyond targeting** that could have implications for compliance with IHL, such as autonomous navigation. Lawand noted that self-navigation has been used in civilian technology such as commercial airplanes, so was not sure what IHL considerations may arise from that particular function.

China and the United States asked questions about the **relationship between self-driving cars and autonomous weapons**, asking what lessons might relevant for the work of the GGE. Bento said there are useful similarities, but the key difference is that while self-driving cars have the potential for violence, they are designed to avoid that violence—whereas LAWS are designed to inflict violence.

Netherlands asked what opportunities there might be for states to **increase transparency around their national weapon reviews**, while Japan asked what panelists thought about the position that these reviews are sufficient. Lawand noted that that

continued on next page



Losing control, continued

the ICRC has been consistently calling for transparency to the extent this will help inform debates about LAWS. In reviewing the legality of these weapons through national weapons reviews, states will identify challenges and maybe solutions, and it is important to share these, to the extent feasible, amongst states in order to learn from each other with a view to setting standards on where limits on autonomy in the use of force should lie.

Russia argued that the general nature of wording in legal weapons reviews is what gives it its advantage. Attempting to agree on a “special approach” vis-a-vis a specific type of weapon could “shatter its universality and entail unnecessary fragmentation”.

In this context, Russia also expressed concern that because states in this GGE process do not yet have a **working definition** of LAWS. Similarly, it argued, there is no common understanding of criteria for the concepts of autonomy or meaningful human control. It is likely, in Russia’s view, that any criteria that could be used here may be applicable to other methods of waging war or weapons development, such as cyber weapons or human enhancement, and questioned why those calling for a prohibition on LAWS were being “selective”.

France argued that **IHL is sufficient** to address LAWS, asking why would we need to develop and codify ethical rules beyond existing law. Algeria later asked if the panelists believe existing law is sufficient.

Parizeau, Lambert, and Obserson disagreed that IHL is sufficient on its own. Parizeau noted that ethics require evaluation that goes beyond legal rules. She argued that behaviour should probably be codified or built into an interactive system to prevent violations of the law or norms, but warned that ethics can’t be programmed. Lambert said that the law is not sufficient and that ethics allow us to underscore the importance of intention. If you look at the principle of proportionality, he explained, it is applicable, but how it is implemented in practice? How does one judge if it is proportional? Interpretation and values come into play—it’s not just a question of executing certain procedures, but of bringing into play knowledge about human dignity.

Russia argued that the problem with IHL is not insufficient legal regulation but lack of compliance to that regulation, which stems from actors ignoring its provisions or lacking national mechanisms for implementation. It thus questioned what the real concern with LAWS is: are we worried about systems with autonomy, or about characteristics such as unpredictability or loss of human control, which in Russia’s view is applicable to other high-tech weapons.

Tuzmukhamedov agreed there is enough law, including IHL, and that compliance with said law is lacking. However, he disagreed that non-compliance is based only on negligence or lack of responsibility. To make the law more effective, he argued, we need to explain it to the recipient, to those who are supposed to apply the law, whether it’s a commander or legal advisor or ultimately the military prosecutor.

Lawand said that as a starting point, existing law is sufficient, and states are indeed first and foremost responsible for applying law to existing weapons systems through national weapon reviews. But she reiterated the ICRC’s point that the novel and unique questions related to LAWS, including about predictability and standards that one applies, may lead to inconsistent results from one review to another.

Thus, she explained, leaving the critical questions raised by new technologies solely to national legal reviews could lead to inconsistent results and issues where perhaps some states will advance with development with minimal limitations while others would impose limitations. The whole point of international discussions and developing common understandings is to resolve and avoid these inconsistencies. While Russia is corrected that there are sensitivities, there are at least some things we could openly discuss, such as standard of predictability.

Switzerland highlighted its working paper on the application of and compliance with existing international law, which proposes a working definition without prejudice for potential future applications. Switzerland sees value in sharing best practices and standards, noting that it feels that “**meaningful human control**” or other similar terms are useful to satisfy valid ethical concerns, ensure compliance with IHL, and respect command and control. The issue now is how much and what kind of human control is necessary. For Switzerland, there are two basic considerations at play: that meaningful human control must be sufficient to ensure the use of LAWS does not produce unlawful outcomes; and dealing with the question of the morality delegating life and death decisions to machines. While there are linkages between these issues, Switzerland wants to treat the legal and ethical questions separately.

On the legal question, the Swiss delegation emphasised, sufficient control over the development and deployment of LAWS must include preparatory measures to implement IHL, and such implementation must be supervised. Switzerland said it is difficult today to conceive of an autonomous weapon system that could reliably operate in full compliance with IHL without human control. It asked the panelists for their views on defining or delineating human control.

continued on next page



Losing control, continued

Lambert argued that if there are limitations to human control over a weapon system, then we are ultimately relinquishing human responsibility. From a legal or ethical angle, he argued, self-learning machines that can self- or –re-programme are removed from what could be considered meaningful human control.

Lawand agreed that IHL assumes there must be meaningful human control (or another relevant concept) over weapon systems. As the factors to be taken into account become more complex, she explained, the level of human supervision and the ability to intervene after the activation of weapon becomes much more important. In this regard, it is critically important that the operator has sufficient information about how the weapon functions and operates and about its foreseeable actions in a given environment—the intersection of capacities and environment in time and space. The greater the complexity of the situation, the greater need for human involvement in the operation to ensure compliance with IHL and to ensure the human remains the decision-maker.

In the context of the discussion on meaningful human control, Bento suggested that perhaps the focus on autonomy is misplaced, as it implies cessation of human control. He argued that we're getting better at developing systems that can learn by themselves, in some cases they are much better at making decisions than humans. In this case, a system might in the future have the capacity to comply with laws and ethics, and that ethics programmed into a machine may even correct certain deficiencies in human ethics or legal rules.

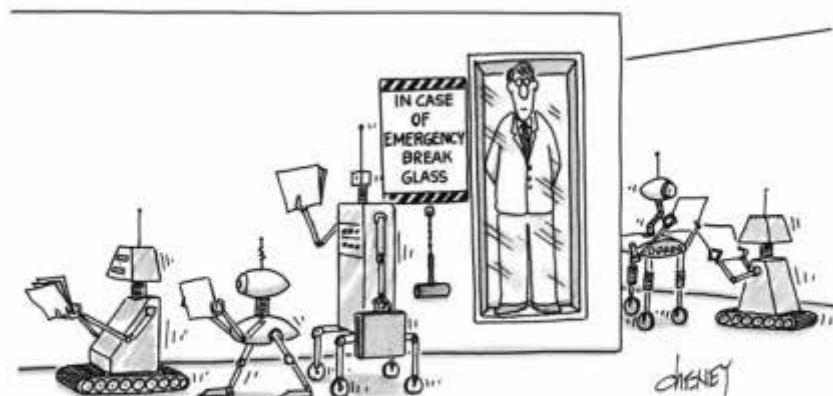
On the other hand, Lambert urged discussants to put human beings at the centre of the debate. Every

time we see humans being deprived of responsibility, he warned, there is a pressing danger for humanity and for law. Once humans are deprived of responsibility, something that truly underpins law is lost. Parizeau similarly argued that the more autonomous the technology, the more need there is to reintroduce the principle of human responsibility in order to address unpredictability.

Drawing on the presentations and discussions over the past two days, Brazil agreed that the capacity for autonomous weapon systems to comply with ethics and morals is not possible. It highlighted the overarching concern from Monday's technology panel that systems that kill and destroy in an automated manner will proceed while the ability of machines to do so in ethical and legal manner will not. **This sets them up for prohibition**, said Brazil, if we want to be positive about curbing imbalance between capacity to kill and destroy and capacity to be moral and ethical.

In response to Brazil, Lawand noted that the ICRC has not yet asked for a new protocol on LAWS, but is not ruling that out. A number of policy and legal pathways are open right now, she noted, and the blinding laser protocol is one of several models that might be available. The ICRC could also imagine a positive requirement of human control over weapon systems, rather than prohibition of certain weapons.

Tuzmukhamedov argued that the protocol on blinding lasers fits description of a system that did exist in some form (at least through anecdotal incidents); and the protocol limited or restricted the use of that system. He left it up to the diplomats to decide whether this a good approach for LAWS. •



CN
COLLECTION

"Meaningful human control"