

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.

## A Critique of the Canberra Guiding Principles on Lethal Autonomous Weapon Systems

<https://www.e-ir.info/2020/04/15/a-critique-of-the-canberra-guiding-principles-on-lethal-autonomous-weapon-systems/>

THOMPSON CHENGETA, APR 15 2020

**This article is part of a series.**

The challenges that are raised by lethal autonomous weapon systems [LAWS] – also known as killer robots – remain pertinent today, even amidst the deadly corona virus pandemic that has taken the world by a storm. Realising that even in times of crisis, it is critical for nations to continue addressing this urgent issue, from 1-2 April 2020, the German Government hosted a Berlin virtual forum on the regulation of LAWS within which 450 people participated from 63 countries. During the forum, the German Government – through its Foreign Minister – reiterated that “letting machines decide over life and death of human beings runs against all of our ethical standards”.

In 2013, before the 2014 ICRC expert meeting referred to by Peter Lee et al, Christof Heyns, the then United Nations (UN) Special Rapporteur on extrajudicial executions, submitted a report to the UN Human Rights Council noting that LAWS raise far-reaching legal, ethical and security concerns. On the legal front are questions whether LAWS are lawful weapons or illegal weapons *per se* and whether they can be used in compliance with international law. Ethicists have questioned whether it is morally acceptable for LAWS to make decisions about who lives or dies. Some scholars have also questioned whether the advent of LAWS will make it too easy for states to use force, thereby undermining international peace and security. As such, not only are “the legal, security and diplomatic communities challenged to assess the impact on the conduct of future warfare brought about by the advent of LAWS” as noted by Peter Lee et al, but they are *also called to determine the legal, moral and ethical acceptability of LAWS in first place*. There is, of course, no consensus on these fundamental questions on LAWS and the way forward.

In a bid to navigate past this stalemate, the Canberra Group suggested the Montreux formula whereby they “set aside the question of LAWS’ legitimacy and desirability to discuss practical options”, namely, provide guiding principles on the use of LAWS. The efforts by the Canberra Group to circumnavigate past the impasse are commendable. Nevertheless, what is the efficacy of the Montreux formula in light of the challenges posed by LAWS? Are the Guiding Principles proposed by the Canberra Group sufficiently speaking to the crux of the LAWS debate? These are the questions I seek to answer below.

### **The Efficacy of the Montreux Formula**

The Canberra Group seeks to “provide guiding principles for the development and use of LAWS without taking a position on the broader political and philosophical questions of acceptability of developing and using autonomous weapons.” As already indicated above, LAWS do not raise only political and philosophical questions. They also raise legal and ethical questions which, in my view, constitute the crux of the matter. The questions on the legality and ethicality of LAWS may not be side-stepped.

Furthermore, the method of avoiding the question of “legitimacy and advisability” as was done in the Montreux Document (herein referred as the “Montreux formula”) may not be very helpful in the case of LAWS. While the question of “legitimacy and advisability” is not necessarily the same as the question of legality or ability/capability to

# A Critique of the Canberra Guiding Principles on Lethal Autonomous Weapon Systems

Written by Thompson Chengeta

comply with international humanitarian law (IHL), there is a dispositive relationship between the two when considering the legality of a weapon.

Where the legality of a weapon is called to question or in cases where a weapon is illegal *per se* – i.e. inherently unlawful by nature – then, its use can never be legitimate or advisable from a legal standpoint. Thus, the question on the legality – and by extension, acceptability of LAWS – cannot be side-stepped or avoided. As such, proceeding to proffer guiding principles on the use of LAWS without addressing whether certain weapon systems may be inherently illegal is tantamount to jumping the hurdle.

Perhaps, the reason why states, in the case of non-state actors and mercenaries, were able to avoid the question of ‘legitimacy and advisability’ is that such non-state actors may be able to comply with IHL regardless of states’ long held positions that the participation of non-state actors/armed groups in war is illegitimate, ill-advised and an offense against the state. Yet, IHL is not concerned with the reasons or legitimacy of war but only that those participating in it comply with weapons law and targeting rules as provided for in IHL and international weapons law.

In short, the Montreux formula on mercenaries can be reconciled with the crux of IHL, namely protection of civilians/protected persons and non-use of prohibited weapons or those that are inherently unlawful. Non-state actors can comply with such rules. The issue is completely different when, like in the case of LAWS, the questions raised include the legality of the weapon itself. It is not an issue that one can shelve and proceed to providing guidelines on the use of the very same weapon whose legality is being challenged or is in question. The questions that are raised by LAWS ought to be answered simultaneously and a piecemeal approach can simply not cut it.

Perhaps, it may be helpful for the Canberra Group to change the approach and title of its proposed guiding principles to “Introducing Guiding Principles to the Use of Autonomy in Weapon Systems”. Like in the case of the ICRC, the Group could underscore the need for internationally agreed limits on autonomy in weapon systems, and like ICRC, acknowledge that certain weapon systems may be unlawful under IHL by virtue of their unbridled autonomy in the critical functions.

## A Critique of the Canberra Guiding Principles

I note that the Canberra Group has indicated that “the principles it sets out represent a starting point for further development.” As such, there is room that some of the concerns I raise below may be addressed at a later stage.

### *Principle 1. IHL Applies to LAWS*

Indeed, it is correct that IHL applies to LAWS. Nevertheless, since other important and relevant branches of international law such as international human rights and laws on responsibility are also applicable to laws, it may be helpful for the principle to refer to the applicability of international law to LAWS instead of just IHL.

More importantly, in the debate on LAWS, debaters often say that international law is applicable to LAWS and emphasise the importance of complying with it. Nevertheless, this emphasis proceeds on the assumption that existing international law is adequate to govern LAWS. Just like the ICRC has noted that LAWS raise questions that go beyond compatibility with our laws to include questions of acceptability to our values, I have explained that existing international law cannot adequately govern LAWS, hence the need for a new legally binding instrument.

Furthermore, in regards to this principle, the Canberra Group notes that “LAWS must be *designed* so they can be operated in a way that carries out commander’s intent”. The fundamental question is what should happen to LAWS that are designed in a way that do not carry out the commander’s intent? Surely, the international community cannot proceed on the faith that no state or entity will develop such LAWS. This is where I emphasise that there is a need for a wholesome approach to the challenges that are posed by LAWS. The Campaign to Stop Killer Robots has provided an example of a treaty that prohibits the kind of weapon systems that the Canberra Group seeks to avoid while indicating useful elements in the lawful use of autonomy in weapon systems. As is clear, the idea with regulation is not at all to have a blanket ban on autonomy in weapon systems. There is a nuanced approach to the problem.

# A Critique of the Canberra Guiding Principles on Lethal Autonomous Weapon Systems

Written by Thompson Chengeta

## *Principle 2. Humans are Responsible for the Employment of LAWS*

The Canberra Group argues that it is erroneous to speak of humans delegating legal or moral responsibility to LAWS because such responsibility **cannot** be delegated to machines and machines **cannot** make ethical or legal choices but “can only function in accordance with their programming”. A scrutiny of the argument made by the Canberra Group in this principle makes it sound as if the authors are saying it looks like a duck, swims like a duck, and quacks like a duck but please don’t call it a duck. Well, it’s duck.

While I agree that machines cannot make ethical choices, I disagree with the notion that responsibility cannot be delegated to machines – even so, wrongly. Where a machine is designed to make all critical decisions without human control, responsibility to make legal and ethical judgments has, in fact, been delegated to the machine. This does not mean that such a delegation is correct, morally or legally acceptable. The fact that machines are incapable of performing a certain task does not mean humans cannot wrongly delegate or burden them with such a task. Therefore, the suggested wording is: “legal or moral responsibility **should not** be delegated to machines since machines **cannot** make legal or ethical choices”. This language is important because LAWS whose design has the net-effect of humans delegating their responsibilities to machines should be prohibited by law.

## *Principle 3. The Principle of Reasonable Foreseeability Applies to LAWS*

In this principle, the Canberra Group refers to the principle of proportionality and precautions in attack. These rules already exist in IHL referenced in principle 1 and constitute what the ICRC has referred to as “existing limits on autonomy”. What is fundamental to note, however, is that in terms of existing IHL, only humans have the responsibility to make judgments on proportionality and continuously take precaution in attack. Now that proportionality entails value judgments, judgments that machines are incapable of making, humans should maintain control of the critical functions. Therefore, the net-effect of these rules is that LAWS whose design exclude human operators from continuously making these assessments and value-judgments stand prohibited by law. It is in this light that I emphasise that it is ineffective to propose principles on the use of weapon systems without articulating or indicating those weapon systems which are prohibited by virtue of their design. This is what a legal instrument on LAWS – like the one suggested by the Campaign to Stop Killer Robots – seeks to do.

## *Principle 4, 5 and 6 on control of LAWS*

The Canberra Group makes a number of indications in principle 4, 5 and 6 relating to control of weapon systems. While the group indicates that autonomous capabilities should enable greater control over desired outcomes, it is not clear how this is achieved in practice. As have been noted by many scholars, certain levels of autonomy in critical functions will not increase control of the desired outcome but lead to unpredictability.

It is also not clear what the Canberra Group means when it says “command and control accountability apply to LAWS”. If it is a matter of law, I have discussed in detail how different modes of responsibility such as individual, command, civil, criminal and state responsibility applies to LAWS. The main point I make is that in terms of these different modes of responsibility, there is a certain level of control that one ought to exercise over a weapon system for them to be legally responsible for all ensuing acts. Such legally required level of control that spells the human-machine dependence in the execution of critical functions outlaws the development of certain weapon systems.

In regards to what levels of control that ought to be exercised over LAWS, the Canberra Group states that control is context dependent. This may need qualification. As has been correctly articulated in the IPRAW report, control of LAWS comes in two modes – control by design and control in use. In regards, to control in use, it is correct to say that control is context dependent. Nevertheless, when it comes to control by design, in particular as it relates to the critical functions of weapon systems, there ought to be a minimum standard of control in the design of weapon systems. The minimum standard of control in the design of weapon systems largely relates to the technical limitations and enablers of limitations on what weapon systems can and cannot do. It is that minimum standard of control that will make it possible for states to determine – through legal review of new weapon systems – whether a particular weapon system is legal or illegal.

# A Critique of the Canberra Guiding Principles on Lethal Autonomous Weapon Systems

Written by Thompson Chengeta

Finally, while the Canberra Group gives reference to ethics, it may be critical to make the issue of ethics more pronounced by including it as a stand-alone principle. While there are disagreements as to which ethics are relevant and how they should apply, there is no doubt that ethics will play a fundamental if not defining role in the international response to LAWS.

## Conclusion

The debate on LAWS is characterised by both converging and diverging point of views. Diverging views notwithstanding, it is clear to stakeholders that the weaponization of artificial intelligence, in particular, autonomy in weapon systems is a very important topic with far reaching consequences. As was indicated by Peter Lee et al, indeed, one of the disagreements centres around how the international community should proceed on the matter.

On the available options and way forward, the Canberra Group indicated that the Guiding Principles are “an imperfect approach in an imperfect world” and “might be the only practical way ahead.” It is important to note, however, that a failure to get an agreement within the framework of the UN Convention on Conventional Weapons (CCW) does not necessarily mean the possibility of a multilateral treaty on the matter is doomed. There are other options, regardless of the associated challenges, for a multilateral treaty outside the UNCCW process.

With the current banphobia that characterises discussions in Geneva, it is important to end this piece by noting that those who seek a regulation, a binding instrument, maintenance of human control over LAWS or a ban seek the same thing. They are not seeking a blanket ban on autonomy. The prohibitions that are sought are nuanced and relate to unbridled autonomy in the critical functions of weapon systems that would result in unpredictability, violations of IHL, human dignity and rights of protected persons. Distinctions that are pointed between a ban, a legal instrument are artificial. Finally, the international community should not desist from making the right laws or ethical decisions because the great powers will not be on board.

---

## About the author:

**Thompson Chengeta** is a European Research Council Fellow on Drone Violence and Artificial Intelligence Ethics at the University of Southampton where he undertakes research and project-related leadership on autonomous weapon systems. His PhD thesis (University of Pretoria) was on international law, ethics and autonomous weapon systems (AWS) while his LLM thesis (Harvard Law School) was on elements that define human control over AWS. Thompson is an executive board member of the Foundation for Responsible Robotics and also serves as an expert member of the International Panel on the Regulation of Autonomous Weapon Systems and the International Committee for Robot Arms Control.