

**European Parliament, Sub-Committee for Security and Defence,
Public Hearing on 'Artificial intelligence and its future impact on security'**

October 10, 2018

Statement of Dr. Marcel DICKOW

Head of International Security division
German Institute for International and Security Affairs, SWP, Berlin

Coordinator of the International Panel on the Regulation of Autonomous Weapons (iPRAW)

**Regulating artificial intelligence in the area of
defence**

Introduction

Madam Chair, Ladies and Gentleman,

First of all let me thank you for giving me the opportunity to speak on the topic of AI and security today. I would like to mention that I speak in my personal capacity as researcher of *Stiftung Wissenschaft und Politik* and not as the coordinator of iPRAW, the International Panel on the Regulation of Autonomous Weapons. Various insights and conclusions of my presentation align with the work and results of iPRAW though. You can find iPRAW's four reports and the forthcoming final report online at ipraw.org.

AI will certainly have an important impact on security and defence applications in the future as we heard earlier today. At the hard security end of the spectrum, AI will be implemented in weapon systems, for example to support autonomous functions such as target identifi-

cation, selection, and engagement/attack. Many more different applications in the field of logistics, combat support, communication, command and control are developed at the moment or will be in the future. It is quite certain that the technology, its application and use will be regulated to a certain extent by militaries itself, states, standardization organizations, alliances under national and international law as it has happened with other conventional technology as well. One reason though stands out: it is the ability of AI-enabled systems to take decisions over the IF, HOW and WHEN to use force, resulting in life or death of humans. This constitutes a fundamentally ethical questions and answering it defines important boundaries for a regulation.

I will focus my presentation today on the topic of the regulation of LAWS (Lethal Autonomous Weapon Systems), as they are called in the framework of the CCW, the Convention on Certain Conventional Weapons, in Geneva. The CCW is not the only international body to discuss LAWS and there is a certain discrepancy between our topic today and the term “AI in weapon systems” on the one hand and LAWS at the other hand. However the CCW is the center of gravity when it comes to legal assessments and possible regulations so far. In a nut-shell this is what you need to know with regard to the discussion process, terminology and legal considerations:

- In 2017 the CCW established a Group of Governmental Experts (GGE) to explore options for a regulation and possible working definitions. Until now it has not made progress with regard to its mandate. Most likely the GGE will continue to debate the topic in 2019.
- So far there is neither an agreed definition on what a LAWS is nor could states parties find common ground whether those weapons already exist or not. Some states parties do have national definitions though.
- The CCW is a forum of international humanitarian law (IHL), more specifically humanitarian arms control. It therefore deals with the legally acceptable use of weapon systems rather than with a regulation of the technology itself.

- The IHL-centered approach has stimulated a discussion whether the L in LAWS is helpful or needed. It has been kept for reasons of continuity.
- It is more accurate to speak of “autonomous functions in weapon systems” rather than of “autonomous weapon systems”, and to avoid to anthropomorphize the machine’s capabilities. For example, iPRAW uses the term “computational systems” to emphasize the importance of the relevant computational techniques that are applied while avoiding expressions such as “a system thinks, decides, etc”. But humans can “delegate decisions and tasks to a machine”.
- As technical definitions are contentious and probably not feasible, the CCW focuses on the human element in controlling the weapon system. The commander and operator are the entities to make the necessary legal judgments according to IHL when it comes to the use of force. In the center of these judgments are the so called “critical functions” of a weapon system: the selection of and engagement with targets.
- The main question therefore is: How does human control over weapon systems need to be designed, implemented and exerted to safeguard compliance with IHL?
- At the heart of IHL-compliance are the following principles and legal judgments: distinction (between combatants and non-combatants), military necessity and proportionality of the use of force, and precaution prior to the engagement. It is imaginable that the notion of “human control” develops into an additional IHL-relevant principle in the future.

Impact of the technology on options for regulation

As I already mentioned, the technology plays an important role for possible regulations, directly and indirectly. In contrast to other weapons-related regulations, variables such as hardware and hardware-centered definitions, numbers of weapons, and weapon types will not suffice to allow for a regulation. It is relatively easy to define a (blinding) laser, to count main battle tanks and to describe a sea-

launched cruise missile. It is nearly impossible to classify a weapon system with autonomous functions in one or multiple steps of the procedures of target selection and engagement. What makes it distinct is the role of the human in every step of the so called “targeting cycle”. The (U.S.) dynamic targeting cycle iPRAW uses throughout its reports consist of 6 steps: find, fix, track, target, engage, and assess (the target/s). It is important to understand that computational methods such as machine learning with neural networks can be used individually or combined in every step of the targeting cycle.

Already today, AI-enabled weapon systems are rather systems of systems than distinct platforms. Its “intelligent”, let’s better call them enhanced, functionalities could be implemented in the war-head itself, but will rather be distributed over a network of sensors, processors, platforms, and effectors. Modular and swarming systems cooperate to fulfill the given tasks, their capabilities develop with their networked multiplicity rather than with a centralized processing unit. Manned-unmanned teaming (MUM-T) adds the human factor to the equation but the human’s role can vary from monitoring to steering a system or parts of it. While it might be easy to tell apart a manned system from a remotely-piloted one, it is almost impossible to verify the role of a human operator in a weapon system with autonomous functions without deep insights into soft- and hardware architecture as well as operational procedures. Autonomous functions are software-based features – when testing and evaluating these capabilities, data and context is as important as code and processing power. With machine learning and other statistical methods, quality and representation of training data turns the balance.

Remaining in control: reasons for regulating LAWS

I see a moral and ethical imperative to remain in control over weapon systems at all times, but reasons for regulating LAWS go far beyond this perspective. Including the ethical dimension, four relevant categories are being discussed in the CCW and in other international fora:

- **Operational:** I quote here from iPRAW's "Focus on the Human-machine Relation in LAWS" report of March 2018:

"At a very pragmatic level there is a strong incentive to prevent fratricide amongst military forces – both personnel and equipment – as a way to maintain combat power and to reduce risk to friendly forces. This overlaps with another pragmatic desire to achieve operational effectiveness, which is based on precision, predictability, and lethal efficiency. At a high, but still pragmatic level, political leaders seek to maintain control over escalation. [...] Militaries utilize a range of mechanisms to achieve and maintain control over violence. Advanced weapon systems have multiple modes, safety mechanisms and highly scripted procedures for use that are developed to lead to successful engagements and to prevent fratricide at the same time. Extensive planning occurs in deliberate and dynamic targeting cycles to ensure the most effective weapons are used in a way that prevents fratricide, achieves operational objectives, and complies with the legal justification and political basis for conducting operations."

In summary: it is a military advantage to have control over the use of one's own weapon systems. Weapon systems need to be predictable and reliable to achieve precision and effectiveness.
- **Legal:** IHL is the legal basis in conflicts and it is universal, even if some states have not ratified specific parts of the regime such as the 1. Additional Protocol to the Geneva Convention. To adhere to the legal principles of IHL requires human judgment before, during, and after the use of force. For technical as well as legal reasons it is not possible to delegate this judgment to machines. Furthermore, an appropriate human judgment requires sufficient situational understanding as a basis for the assessment. Only the involvement of humans in the targeting process and the ability of the operator/commander to intervene during operation safeguards responsibility and accountability within the military hierarchy.

- **International security and stability:** Already today the technology to achieve autonomous functions in weapon systems proliferates intensely. One reason is the importance of civilian drivers and stakeholders (science, industry, states) in its development – something we all support and do not want to regulate per se. – In addition a growing number of states is interested in harvesting potential military advantages of the respective technology, such as higher precision and increased fighting speed. An increasing number of countries have developed security-related AI-strategies and have set in place military development and procurement programs in the field of remotely-piloted and autonomous systems. This highlights the trend towards AI in defence. We observe arms dynamics in the field of LAWS for example amongst the P5 States, which are, at the same time, reluctant towards a regulative approach in the CCW. Their interest in “fighting at machine speed” stimulates a risk of crisis escalation and regional destabilization. There will be an inevitable trade-off for increased battlefield speed: the loss of operational control. This is even more true when adversary LAWS interact in the future.
- **Ethical:** Is it in principle unethical to delegate life and death decisions to machines? In general and specifically in the area of defence? The answer to this question depends on the ethical theory you apply. To this end the last iPRAW report on “Ethical Implications for a Regulation of LAWS” investigates the role of human dignity as a key concept in ethics. It specifically asks two questions to explore the consequences for a possible regulation:
 1. Does autonomy in weapon systems break the link to moral agency [and therefore violates the dignity of the victim/target of an attack (added by the author)]?
 2. Can consequences (the expected advantage of a specific application of force in combination with autonomous functions in weapon systems) override issues of human dignity?

By addressing these basic questions iPRAW comes to the following conclusion: “Depending on the moral position, one would assume or deny that autonomous functions in weapon systems break the link to moral agency. In the first case, it would be necessary to safeguard moral agency through human control, in the latter case one would want to safeguard the ability to use a weapon system lawfully at the current state of technology. In consequence, both positions would require human control in both the design of the system and its use. Inherent in both views is an acknowledgment – tacit or explicit – of the principle of human control.”

Conclusion

Let me emphasize that there are important incentives and objectives to regulate LAWS. As a direct regulation of the technology is probably unfeasible due to the lack of definitions of the subject (LAWS), the CCW discusses – very much in line with its mandate – the regulation of the *use* of these weapon systems. Stronger regulations could also include the development, testing and fielding of respective systems. The majority of the scientific community, including iPRAW, civil society, and a growing number of states call for a positive regulation by requesting human control as the underlying principle for the design and use of such weapon systems. Again iPRAW has proposed two criteria to specify and safeguard the concept of human control: **control by design** and **in use** on the one hand, **situational understanding** and the **ability to intervene** for the operator/commander of such systems on the other hand. You will find the details in iPRAW’s 3rd report.

Weapons reviews (in accordance with Article 36 of the 1. Additional Protocol to the Geneva Convention), meaning testing and evaluation of weapon systems before fielding them, are a necessary but not sufficient prerequisite to maintain human control over weapon systems during operation. Moral agency and human judgment with regard

and prior to the use of force must not be delegated to machines and cannot be substituted by built-in technical procedures and features.

Let me again stress the benefits of AI in defence applications in general. However they must not lead to a lack or loss of human control when it comes to the use of force. In logistics, intelligence, and other fields, AI can play an important role to increase efficiency, like in any civilian field as well. Necessary condition for a responsible use, such as transparency and accountability with regard to algorithms and data certainly do apply. Even in weapon systems autonomous functions can enhance precision, for example in highly automated defensive systems against incoming munition. However computational methods such as (unsupervised) machine learning substantially challenges the concept of human control in weapon systems. Design and operational procedures (including training of personal) must balance and limit such techniques to a degree conceivable in real time for human operators and commanders.

I'm happy to answer your questions in the following and thank you for your attention.