



Security Policy Working Paper No. 33/2018

Legal Regulations for Autonomous Weapon Systems: Potentials and Limitations

by the authors of the “Young Leaders in Security Policy” working group

Research and development for autonomous weapon systems has in recent years gone up exponentially. Although derided as mere science fiction (“killer robots”) only a few years ago, the employment of autonomy in weapon systems has in some cases already become a reality – for instance, in the Bundeswehr’s PATRIOT and MANTIS air defence systems. The development of (semi-)autonomous weapon systems inevitably reduces or, in extreme cases, even completely eliminates human control. That is cause for concern on several levels. For instance, there are fears that largely autonomous weapon systems without sufficient human control might lower the inhibition threshold in military conflicts, considerably accelerate warfare, increase the number of casualties, drive up the risk of escalation, and make compliance with international humanitarian law more difficult. What precautions can be taken against undesirable outcomes?

Regulatory hurdle: What should be the object of regulation?

In light of the possible risks and challenges of (semi-)autonomous weapon systems, a number of countries have already begun a serious debate on regulation; for instance, in the framework of a working group on the UN Convention on Certain Conventional Weapons. A comprehensive and global solution is, however, still a long way off. A project of this nature already faces a headwind when it attempts to exactly define the object and scope of the intended regulation. So what exactly does the term “autonomous” cover with regard to weapon systems? Should the development of such systems already be restricted or is it enough to simply regulate their employment? If the objective is a comprehensive set of regulations, it would make sense to develop internationally binding rules already for the development of autonomous weapon systems to prevent them becoming available in the first place. Such an approach, however, would require a clearly defined object as a concrete point of reference for the regulation.

It becomes apparent how difficult it is to use the blanket term “autonomous weapon systems” to describe this object. For there are indeed several degrees of autonomy and different ways in which to apply it in weapon systems. Only very few systems act fully autonomously, and not all semi-autonomous or autonomy-based systems are cause for concern. For instance, a number of defensive weapons, particularly those against rockets, cruise missiles and artillery shells on land and at sea, employ systems that independently detect and engage incoming airborne targets. Due to longer human response times and the speed of the inbound missiles, humans would be unable to carry out these functions. However, these types of systems are not particularly worrying, as they are used in a narrowly defined context for purely defensive purposes and shooting down unmanned missiles does not cause human casualties.

Autonomous weapon systems have come in for criticism mainly because of ideas of how they might be employed in the future: as *offensive weapon systems operating with full autonomy* that themselves make and execute attack and/or kill decisions – for instance, drones that fly into a target area and, depending on the situation, engage targets from the air on their own without waiting to receive orders from a human decision-maker. The phenomenon of autonomy is thus not in itself a deciding factor, but rather the question of how and for what purpose this autonomy is employed. That is precisely why it is so much harder to pin down what the concrete object of any development-focused regulation would be.

It would accordingly be more appropriate to focus more on the type and scope of “autonomy *in* weapon systems” rather than employing the term “autonomous weapon systems” as a catch-all without sufficient distinction. The decisive question is then to determine *which forms* of autonomy are grounds for concern (or not). If the development of certain “problematic” systems or system functions were to be restricted, the concrete way in which they are “problematic” would have to be described. This is hard to do in an abstract manner, however. Any international, treaty-based regulation that tries to restrict autonomy in weapon systems in a generalized sense would thus already run into choppy waters when it comes to defining an object that can be regulated by an agreement and would in all probability not receive much support in the international community. This is why regulation seems hard to do. Seen from this perspective, a more efficient approach appears to be discussing the *employment* of (semi-)autonomous weapon systems and the problems surrounding the various ways in which they could be used.

Opportunities and problems of regulating the employment of (semi-)autonomous weapon systems

In light of the concerns laid out in this paper, focusing primarily on regulating the *employment* of (semi-)autonomous weapon systems still raises the question as to which criteria should be used to distinguish between “problematic” and “non-problematic” variations of autonomy in weapon systems – albeit this time, the focus is on the mode of employment, while development takes a back seat. The already existing rules of international humanitarian law can be used to develop a basic principle for what constitutes a “problematic” mode of employment. In this context, the concern is often voiced that with increasing autonomy, complying with certain central tenets of international humanitarian law will become impossible – these being primarily the principles of distinction between combatants and civilians, the identification of admissible targets, and the prohibition of disproportionate collateral damage. International humanitarian law, however, developed in light of conventional weapon systems. Yet precisely these conventional weapons do *not* act autonomously in line with rules that they themselves have set, but rather are merely tools employed by a user who makes the decisions. Soldiers who operate them are as individuals responsible for their actions on operations; they, their commanders and their political leadership are held to account.

The core difficulty of any largely autonomous system, on the other hand, is that it would, to a certain extent at least, be programmed to itself formulate if-then statements without being in any way held accountable. These statements would merely be the result of a calculation process that, though based on programming done by a human, leads to a concrete result unforeseeable at the time of programming. The required chain of accountability is thus broken, so that such a system can indeed be considered “problematic” even outside the scope of conventional international humanitarian law. Unlike soldiers, who know the reasoning that led to their decision, can describe and justify it and are thus accountable for their own compliance with international law, employing (semi-)autonomous weapon systems risks creating a responsibility gap, as the reasoning process that led to the decision can then be reconstructed only based on the outcome of that decision. This precisely is the main point of criticism to be lodged against particular autonomous weapon system.

As it not in any country's national interest to artificially create such a responsibility gap, however, it seems a realistic endeavour to persuade governments to support an international treaty that is more narrow in scope and that closes this responsibility gap between the decision by a human decision-maker to employ a weapon system and the moment in which the weapon system launches an attack. An international treaty could here be worded in such a way as to allow the employment of autonomous weapon systems only if it is ensured that there is "*command responsibility*" every time the weapon system is employed, i.e., a human supervisor is in charge and accountable for every employment. If this supervisor finds that the employment of an autonomous weapon system cannot guarantee compliance with international humanitarian law to the same extent as using human actors can, the system must either not be employed, or the liability regulations of the law of armed conflict and international criminal law apply along with the general rules of state responsibility. Employing this relatively simple mechanism, which underwrites the rules of international humanitarian law, could additionally, from the development stage, help place the focus on those systems that are legally permissible from the outset.

Opportunities and problems of regulating the development of (semi-)autonomous weapon systems

However, such a regulatory instrument aimed primarily at the employment of these systems has considerable shortcomings. For instance, it is doubtful whether the outlined approach can achieve more than simply untangling questions of ex-post responsibility. In addition, at least, the aim should be to establish a set of international, legally binding rules that, ex ante, prevent the development of autonomous weapon systems that are designed in such a way as to almost inevitably violate international law. Focusing exclusively on the *employment* of autonomous weapon systems is not sufficient. Rather, an international regulatory tool should be put in place in order to limit and monitor the development of such weapons.

A central problem of weapon systems that act largely autonomously is that in a number of typical, ethically and legally dicey situations, even the forward march of technology cannot replace human judgement. It would be a misapprehension to conclude that autonomous weapons can at some point be programmed accurately enough to distinguish between legitimate and illegitimate targets more precisely than a human could. This becomes particularly evident when possible concrete fields of application and examples are considered. In many conflict environments, it is hard to distinguish combatants and civilians from one another, as nobody is uniform. Civilians are used as human shields, and combatants can always surrender to their enemies and thus no longer be considered a legitimate target. Prisoners of war are under special protection, but cannot always be easily distinguished from combatants. In addition, the approach that prioritises the employment of autonomous weapon systems focusses primarily on "traditional" conflicts between nation-states. However, most current armed conflicts are not carried out between states but within states, mostly against and by using irregular armed groups and militias. In these conflict situations, even human actors do not usually clearly distinguish between civilians and combatants.

Thus it is only in the rarest of cases that making the right decision can be reduced to a mathematical equation. Rather, it is the result of a complex process of assessing and weighing options. It is not about combining information in line with a fixed scheme, but about making highly complex, case-by-case decisions. The myriad of facts that have to be assessed for such decisions cannot easily be captured by an algorithm. Serious mistakes would, in the truest sense, be programmed in advance. This in turn means that from development, autonomy in weapon systems has to be designed in such a way that the final decision to kill is in the hands of humans who are equipped with the practical sense of judgement required in these circumstances ("meaningful human control").

In the debate around regulating autonomous weapons systems, the possible proliferation of such weapons must be considered in addition to the decision-making determining their employment. Systems of this nature do not have to be complex, expensive and “smart”. They can potentially also be simple, affordable and “dumb”. They could consequently be fielded by regular armies as well as by irregular armed groups or, due to their autonomous nature, even replace an irregular armed group. It is very easy to imagine governments or warlords in a civil war equipping themselves with cheap autonomous weapons in order to increase their chances of winning, provided there is a corresponding market. Autonomous weapon systems can be employed in a fashion similar to militias; they similarly complicate the establishment of direct causal links and are free from human scruples, for example when executing a so-called “cleansing”, and are much more cost-effective in the long term. Potentially, this could mean a sort of “minimal-effort genocide”. Similar considerations apply to the dangers of an employment of largely autonomous systems by terrorist groups.

The problems only roughly sketched out in this paper demonstrate that including the development stage of autonomous weapons in an international set of regulations is important particularly for preventing the potential misuse of autonomous weapon systems. It would be sheer negligence to not also consider future, unexpected consequences of the development of new weapons or weapon types. The first step to controlling, from the very outset, the misuse and proliferation of autonomous weapons is to regulate and limit their development. This could be achieved, for instance, by placing internationally binding restrictions on the technological implementation of autonomy in specific weapon systems. The aim should be to consider misuse and proliferation not simply as an inevitable adverse effect, the consequences of which can be mitigated only after the fact by an ex-post ban of certain autonomous weapons. Instead, to pre-empt misuse as much as possible, rules should be established in a precautionary manner.

Concluding remarks

At this point, the international community has an historic opportunity to reflect not only on the advantages, but also and most importantly on the possible and often not intended negative effects of a new, highly controversial and potentially threatening type of weapon at a fairly early stage of its development. To this effect, international negotiations should examine all types of international agreements and be open to every dimension of the debate. If from the start, regulation is limited to a minimum consensus, we risk missing this unique opportunity of putting a stop to potentially undesirable developments before they come about.

The “Young Leaders in Security Policy” working group was founded in April 2015 by the Federal Academy for Security Policy and its Association of Friends. One of the working group’s goals is to encourage young professionals from politics, sciences, public administration, business, churches and the armed forces to share their ideas and views on security policy issues.