Federal Academy
for Security Policy

Security Policy Working Paper No. 26/2018

Artificial Intelligence in the Armed Forces
On the need for regulation regarding autonomy in weapon systems

by Frank Sauer

Armed forces all over the world have started to explore the use of artificial intelligence (AI) for military purposes. The Bundeswehr is no exception. Spheres in which AI can demonstrate its strengths as well as optimize and accelerate processes in the Bundeswehr include not only logistics, predictive maintenance, combat service support and operational command and control, but also the analysis of large databases to spot developing crises at an early stage. An area that is considerably more sensitive is the use of AI to increase "autonomy" in weapon systems – to the point of fully autonomous weapon systems that can, without being controlled by a human, select and engage targets. This working paper looks into the functional character of weapon autonomy and examines which of its aspects are actually new. It then outlines the risks inherent in the use of fully autonomous weapon systems. Finally, this paper presents three proposals on how Germany should position itself internationally and nationally with regard to the Bundeswehr while avoiding risks and seizing opportunities.

Scientists and representatives of civilian technology companies, the drivers of innovation in the field of AI, have been warning about a paradigm shift in warfare for years, for which they have received considerable media attention. One of these was the late Stephen Hawking, along with others such as Elon Musk or Demis Hassabis and Mustafa Suleyman, the founders of Google's AI development company DeepMind. They all draw attention to the risks with regard to international law, security policy and ethics if future weapon systems should operate increasingly outside of meaningful human control. At the level of the United Nations (UN), these warnings from civil society have promoted multilateral talks about possible arms control for fully autonomous weapon systems, for which the UN term is LAWS (Lethal Autonomous Weapon Systems). These talks have been ongoing since 2014 within the framework of the UN Convention on Certain Conventional Weapons (CCW) in Geneva.

Weapon autonomy as a functionality

So far, there exists no generally accepted (let alone internationally agreed) definition of autonomy in weapon systems. However, in the international expert discourse, a functional definition of autonomy is increasingly finding acceptance – not least due to the fact that important actors such as the United States or the International Committee of the Red Cross (ICRC) have, in policy documents and statements, officially adopted such an understanding of autonomy in weapons. According to this definition, a weapon is considered fully autonomous if it completes the decision cycle for target engagement on its own, that is, after activation controlled solely by its software and without any human control or supervision, unlike a remotely controlled system. This targeting cycle includes the stages *find, fix, track, target, engage* and *assess* (abbreviated as F2T2EA). Fully autonomous weapons would thus be beyond human control when selecting and engaging targets – actions that are especially sensitive from a legal, ethical and political point of view and that are regarded by the ICRC as the "critical functions" of the targeting cycle.

 | ISSN 2366-0805

If this understanding is applied, some existing weapon systems, on closer examination, already qualify as fully autonomous, or, in UN parlance, as examples of already existing LAWS. For instance, the Israeli system *Harpy* is directed against radar installations and, for this limited purpose at least (circling over an area and engaging enemy air defenses), it already goes through the targeting cycle without human control. Among experts, *Harpy* is thus unanimously deemed an example of an existing LAWS. Therefore, the LAWS debate is by no means about "weapons of the future."

Defense systems (of the past) that engage targets "only reactively" are often contrasted with "autonomous" systems (of the future) that have offensive capabilities. However, this differentiation also does not stand up to closer examination in light of a functional definition of autonomy. It is certainly true that for decades, defense systems against missiles, artillery shells or mortar bombs have been employed to – under extreme time pressure and, if necessary, without human intervention – engage incoming projectiles. Unlike such weapons mounted on ships, for instance, or emplaced on the perimeter around military bases, LAWS are often considered mobile weapon systems, or rather weapon platforms, that introduce modern technologies such as machine learning into military operations.[1]

However, autonomy does *not* depend on the advanced nature of the employed technologies. Several long-standing defense systems – including, for example, the PATRIOT surface-to-air missile defense system that is also used by the Bundeswehr – are generally capable of acting autonomously as regards the critical functions of target selection and engagement. They are, consequently, "fully autonomous" according to the functional perspective adopted here. To put it bluntly, functional autonomy in weapons is not as new as people make it out to be, it is just that nobody really took notice up until recently. The implications are becoming more apparent only now because full autonomy is increasingly being used in not only defensive weapons but other, mobile systems. Just as autonomy is not bound to specific high technology, for that matter, it is also not necessarily tied to single specific weapon systems. The "intelligence" enabling autonomy can also be spread out across interconnected weapon systems or a swarm.

Interim conclusion: Autonomy in weapons is not categorically distinctive

Autonomy in weapon systems thus characterizes neither a specific nor a revolutionary new *category* of weapons. The idea that a categorical distinction can be made between "autonomous weapon systems" on the one hand and "non-autonomous weapon systems" on the other is misleading. Not least because the new technical possibilities that are currently being introduced with AI can be used to retrofit old systems ("autonomizing" them, so to speak), and because new remotely controlled systems will almost certainly be optionally autonomous to varying degrees. In short: For future weapon systems, it will not be apparent "how autonomous" they are or can be. Insisting on a definition of LAWS that is based on categorically distinctive features and an attempt to distinguish between "autonomous vs. non-autonomous" not only does not live up to the subject matter intellectually, it has also long since become a hindrance for the arms control talks at the UN's CCW in Geneva, allowing for a lot of heel-dragging.

In contrast, it is more intellectually coherent and also makes more sense politically to follow the example of the US and the ICRC and accept a functional definition of autonomy, even if some may find awarding comparably "old" weapon systems full autonomy somewhat counterintuitive. It would be wrong, though, to then conclude from this that established defense systems, which have so far been uncontroversial (and rightly so), have an autonomy problem all of a sudden. The practice of autonomously defending against shells and missiles has never previously posed any significant problems with regard to international law, ethics, and security policy, and it continues to be unproblematic in this respect. In fact, the protective function of these systems when defending against inanimate targets is most welcome, and in view of the development of hypervelocity missiles, the military requirement for this manner of defense will only increase.

[1] Most progress in the field of AI currently stems from machine learning.

Therefore, political decision-makers should see the LAWS debate not as one about technology, but rather about specific, acceptable or not acceptable military *practices* when employing weapon systems with autonomous functions. The key question to ask is thus when and to what extent human control is reduced or completely replaced when using a weapon system – and what the respective consequences may be. So new technologies from the field of AI certainly do play an enabling role in LAWS development, but the real novelty lies not in technology but in practice, that is, in the military application of autonomous functionality in weapon systems in areas in which autonomy has so far not been present on the battlefield. This leads to scenarios in which tanks, ships, aircraft or infantry might in the future be engaged in a fully autonomous fashion. And in contrast to the autonomous defense against incoming rounds, i.e., inanimate objects, full autonomy in weapons in such a context would give rise to extremely problematic consequences indeed.

Autonomous weapons and the risks

A number of serious problems would arise if autonomy in weapons, which so far is used only in isolated cases such as Harpy or restricted to defensive systems such as PATRIOT, is also used in mobile systems operating without meaningful human control and executing critical functions to engage inhabited targets or people. From an International Humanitarian Law (IHL) point of view, this practice would cause a responsibility gap. For it is not clear who would be responsible should such systems cause civilian suffering that is disproportionate to the military value of the engagement, unjustifiable and therefore illegal. The currently emerging global arms race around full autonomy in weapon systems also threatens global military stability and increases the risk of unintended escalation. The financial markets already demonstrate how unpredictable algorithm interactions can sometimes result in so-called *flash crashes*. Without human control acting as a fail-safe and deceleration mechanism, a *flash war* triggered by machines is a real danger. The final, most important objection to an unregulated use of full autonomy in weapon systems is a moral one, namely that it violates human dignity to delegate decisions about life and death on the battlefield to algorithms. For by outsourcing the killing in war to machines and having them automatically "work through" their tasks, human beings are reduced to mere data points and turned into objects. The dead may not much care whether their death was caused by an algorithm or another human being. But the society that allows such a practice and no longer troubles its collective human conscience with war-time killing risks nothing less than giving up the most basic values of civilization and fundamental principles of humanity.

Conclusion: On the need for regulatory action and a nuanced use of autonomy in weapons

The course for the future use of autonomy in weapon systems is being set today. Particularly with regard to the critical functions of weapon systems, the correct path to choose in light of the risks outlined above is to generally hold on to human control – the prominent exception being the above-mentioned defense systems against shells and missiles. In other words, a nuanced, careful and prudent approach is required with regard to the military use of AI. In light of this realization, Germany is playing an active role in the arms control talks at the UN CCW in Geneva where the international community is discussing multilateral arms control regarding LAWS. Within the CCW, 26 countries are currently demanding that weapons operating without human control be prohibited under international law. Austria has committed itself to this demand and has been very active in supporting it during the most recent talks in August 2018. China has been sending similar but still somewhat unclear signals since April 2018. However, the US and Russia in particular are strongly opposed to any kind of regulation.

Germany has so far not joined the group of ban advocates at the CCW, but in its coalition agreements of 2013 and 2018,[2] it unequivocally rejects fully autonomous weapon systems and demands an international ban. Together with France, Germany is now trying to work towards that goal by way of a political declaration and other "soft law" instruments, staking out a middle ground between the supporters and the opponents of a ban, hoping to gradually move the process forward this way. At the national level, high-ranking

[2]"We reject autonomous weapon systems that are not under human control. We seek a global ban on them." (7027 to 7028)

members of the German military have reaffirmed the position of the Federal Government, as the Chief of the German Cyber and Information Domain Service, Lieutenant General Ludwig Leinhos, did at the 2018 Munich Security Conference. Regarding the need for national regulation, it is not enough for the Bundeswehr to simply state that the provisions of IHL will always be observed, even in light of progressing autonomy in weapon systems. This stance does not do justice to the far-reaching implications of this development – an emerging new capability and possibly a redefining of the relationship between humans and machines in war. On the contrary, new regulation is needed.

The targets formulated in the government's coalition agreement could be implemented or pursued nationally and internationally in three steps:

Step 1) Draw up a policy directive for autonomy in weapon systems
This document should adopt a functional understanding of autonomy based on the critical functions of target selection and target engagement. It would establish a set of forward-looking regulations for maintaining meaningful human control over all weapon systems of the Bundeswehr. Specifically, it would prescribe a human-machine interaction in which the maximum level of autonomy permitted would be for the software to select targets, which would then have to be approved by a human operator prior to any engagement. A higher degree of machine autonomy would be acceptable only for current and future defensive systems, as explained in step 2.[3] This document could be reviewed every five years in light of technological developments and updated if required.

Step 2) Continued, regulated use of fully autonomous defense systems
The policy directive should declare the continued use of fully autonomous defensive systems to be an exception, but proactively regulate such a use by means of the following framework conditions and restrictions regarding design, construction, and operation. In line with these conditions, fully autonomous defense systems of the Bundeswehr may

- be employed only under time pressure and defensively against direct attacks by inanimate military objects;
- repeatedly execute independent of external influences or control only a small number of precisely defined preprogrammed actions, the effects of which must be highly predictable;
- be used only in uncluttered and easy-to-assess environments with a very low risk of harming civilians;
- operate only if they are permanently installed, for example, on board a ship or in a military camp, but not as independently mobile weapon systems.

Step 3) Intensify German efforts towards establishing a CCW protocol that bans LAWS in a manner that is verifiable and binding under international law
As the positions on either side of the debate in the CCW's Group of Governmental Experts (GGE) grow increasingly entrenched, the course pursued by Germany and France in Geneva currently provides an alternative both to an immediate ban (considered premature by some) and a complete standstill. However, the German-French middle course also bears a risk. Its first step, a non-binding political declaration, could, if intentionally misinterpreted as a *last* step by third parties, curtail talks long before Germany's goal of a ban, as set forward in the coalition agreement, is achieved. In 2019, Germany should therefore state this goal more clearly and, additionally, press for a negotiating mandate for the GGE. That way, the talks at the UN CCW in Geneva could at least, finally, have a chance of turning into real negotiations.

Dr. Frank Sauer researches and teaches at Bundeswehr University Munich. His Twitter account is @drfranksauer. This article reflects his personal opinion.

[3] For other legacy systems of the Bundeswehr that may fall within a functional definition of weapon autonomy, such as certain sea mines, possible exceptions would have to be considered on a case-by-case basis.