Metis

Study

The security-policy effects of digitisation: Future forms of conflict and conflict management

No. 01 | February 2018

Summary

The autonomy of machines is the most important future trend in the digital age. With regard to weapon systems, autonomy promises to act as a force multiplier and to permit higher operational speeds and more precise effects. However, in matters of conflict management, the elimination of human control carries operational risks of a legal and ethical nature as well as strategic risks due to new, escalation-prone forms of conflict. As regards a Bundeswehr use of autonomous weapon systems, this report thus recommends a nuanced approach that helps the Bundeswehr exploit potential advantages while minimising the risks. On a national level, this approach includes preparing a policy document to make sure that the selection and engagement of targets always remains under meaningful human control except for defensive systems. On an international level, it encompasses efforts toward an international, verifiable set of regulations on autonomy in weapon systems that are legally binding under international law.

Topic outline: Digitisation and autonomy

With the progression of the Information Age and the process of digitisation, the scope and complexity of tasks that humankind delegates to computers and machines is growing. Robotics and artificial intelligence (AI), in particular, currently represent key technologies in this development. By now, the general public has become aware of the increasing importance of algorithms and machine "autonomy" in the form of Apple's digital assistant Siri or Tesla's Autopilot driver assistance system.[1]

The importance of this development is simultaneously over- and underestimated. It is overestimated because the "intelligence" component of the term artificial intelligence evokes the wrong kind of association.[2] Although machine learning (ML), the process that is mainly responsible for recent successes in the field of AI, is a powerful tool for pattern recognition (for example in images and written or spoken text),[3] it is limited to narrowly formulated tasks. It thus cannot be compared to the flexible and generalised skills associated with human intelligence. At the same time, the impact such "unintelligent" systems may nevertheless have is underestimated. For instance, algorithm-generated and easily-to-manipulate "filter bubbles" in social networks, which currently interfere with the function of the free media as the fourth estate, impede established processes of democratically shaping public opinion.

The coming phase of digitisation thus already raises pressing social questions for the future. As part of "Industry 4.0", these questions also include economic and social as well as (machine-) ethical issues, for example the use of robots in nursing care services.

1 This study focuses on the security-policy implications of autonomous machines as an aspect of digital progress that has implications for the future. Issues related to cyberspace are not discussed.

2 The broadly and not uniformly defined concept of artificial intelligence comprises a variety of different software-based techniques and procedures for task automation that until now required the application of human intelligence. That is why in the following, the rather unhelpful term "AI" is generally avoided. Instead, the text refers to the concrete techniques, such as automated image recognition, that are relevant to a particular discussion.

3 Currently, automated image recognition still requires a lot of computing power and huge amounts of data to allow deep learning (with neural networks for data classification) to work. The process is expected to become significantly more efficient and less complicated in the near future.

Machine autonomy in the armed forces

The shift in task division between human and machine also represents new advantages and challenges for the armed forces. The focus of this study, however, is not on already existing applications such as those used for the automated fusion and analysis of data or in C2 support and battle management systems. Instead, it takes a future-oriented look at the phenomenon of increasing autonomy in weapon systems (AWS).[4]

For a weapon system to be fully autonomous as defined according to the US AWS policy directive and the International Committee of the Red Cross (ICRC), it must, after activation, with the help of sensors and software, be able to go through an entire targeting cycle autonomously, that is, without any human control or supervision, unlike remote-controlled systems. The system actively seeks and finds targets, fixes and tracks them and also performs, without human intervention, the "critical" functions of target selection and engagement (find, fix, track, select, engage, assess).

In principle, weapon systems that autonomously engage targets in this way are not new. Defence systems such as PATRIOT have been used on operations for decades.[5] Under time pressure, these systems can also engage targets without human intervention (terminal defence). However, they are usually stationary, perform the same pre-programmed actions over and over again, and are directed against ordinance, in other words against inanimate targets.

This study, on the other hand, is interested in the autonomy of mobile systems that operate in dynamic, unstructured, open environments over longer periods of time and that require no human intervention to carry out the targeting cycle. Autonomous weapons meeting this definition have, albeit with narrow tasks, already been fielded. One example is the Israeli anti-radar loitering munition Harpy (in the case of Harpy, the scope of application is limited to cruising over an area and engaging enemy air-defence radar systems).

Weapon autonomy: Advantages and challenges

What are the advantages and challenges of AWS developments for Germany and the Bundeswehr at an operational and strategic level?

Generally speaking, the advantage of autonomous weapon systems is that they can carry out monotonous, unpleasant and, especially, dangerous tasks. In a more concrete sense, three other advantages are currently being debated.

Firstly, AWS serves as a force multiplier: In the future, one single soldier will control a large number of autonomous systems or swarms.

Secondly, autonomy makes control and communications links optional. These links are susceptible to disruption as well as being vulnerable to hijacking, and they may occasionally give away the location of a system. In addition, there is always a time delay between human remote-control commands and their execution. Autonomous systems that are not dependent on these links promise no-latency – and therefore higher – operational speed and operational capability in situations in which control and communication are difficult, impossible or undesirable due to stealth requirements.

Thirdly, the fusion of real-time reconnaissance, decision-making speed and a precise use of weapons without time delay could make it easier to observe the rules of war by avoiding civilian casualties and damage to civilian objects.

However, this third point in particular is highly controversial among legal professionals and AI/robotics experts. The question of whether autonomous weapon systems make it easier or rather harder to carry out military operations that conform to international law leads us to the implications of AWS in matters of conflict management.

Future conflict management: Will humans still be in control?

Conflict management in this case is understood in an operational sense, that is, in terms of military operations. AWS do, in fact, introduce a new factor into the equation. For it should be noted: Premature comparisons with existing homing munitions or projectiles that follow a ballistic trajectory are misleading. AWS are not equivalent to sending munitions along a trajectory that can no longer be influenced or beyond a certain recovery point. This is because the idea of autonomy is – to put it simply – to allow a "mission" to be assigned to a weapon system. The system will then operate without human control and supervision, possibly over a lengthy period of time, and it will make its own "decisions" regarding target engagement. Accordingly, this represents a new situation (otherwise, the development would not even be pursued in the first place). Reassessment and independent legal evaluation are thus required. A number of problems arise.

Firstly, the question as to whether a legal, autonomous completion of the targeting cycle is even technically feasible, i.e. whether AWS is even able to conform with international law, remains unanswered. According to the current state of the art, at least, this question must be

4 The acronym LAWS (for "lethal autonomous weapon systems") is often used as well, especially in the context of the current debate at the United Nations Convention on Certain Conventional Weapons in Geneva.

5 Mines can be also included in this broad definition of autonomous weapons – at least the ones that perform a type of target selection, using certain signatures, that is not based merely on a primitive victim-activated on/off-mechanism with no self-regulating loop.

answered in the negative. Few if any representatives from the relevant technical fields believe it to be possible for a machine to discriminate between legal and non-legal targets (e.g. between combatants and civilians – which can be extremely difficult, even for humans, because it depends on context and an understanding of social meaning) and make assessments as to the appropriateness of military means. Moreover, pattern recognition systems based on deep neural networks, which represent the current state of the art in the field of automated image recognition, have proved to be extremely susceptible to manipulation.[6]

Secondly, unlike with the firing of a homing missile, when it comes to the use of AWS, there is currently no legal majority opinion concerning the question of where and/or with whom the responsibility for the employment of weapons in the context of the international law of war lies. Even assuming that autonomous systems operate without errors, this would pose a problem. Since errors caused by software and hardware or the fog of war as well as enemy influence are unavoidable, AWS carry the risk of creating an inacceptable legal "responsibility gap" in case of legal violations.

Thirdly, AWS also have implication for the nexus between international law and ethics. For if the systems are not able to, for instance, reliably identify persons who are no longer engaged in combat activities due to injury or because they have surrendered (are "hors de combat" under international law) and treat them appropriately, then the implications of this go beyond a purely legal context and individual cases. Here, the development of weapons autonomy comes into conflict with the general obligation to not deprive humans of their dignity and expose them to unnecessary cruelty, even in war.

Autonomy in weapon systems that includes the critical functions of target selection and target engagement, except for defence against incoming munitions, therefore carries operational risks with regard to ethics and the international law of war, due to the fact that human control (and therefore legal and ethical responsibility) has been replaced. From the point of view of Germany and the Bundeswehr, these risks must be avoided. The Bundeswehr should not delegate the decision to kill to machines whose conformity with the international law of war is debatable and that do not understand ethics or the difference between life and death.

Future forms of conflict: Battlefield singularity and escalation dynamics?

The strategic impact of autonomy on forms of conflict at a global level is largely influenced by the aforementioned increase in operational speed.

Autonomy begets autonomy because speed, which is defined as the ability to act before an adversary has completed his decision cycle, promises key advantages. However, as Robert Work, the US Deputy Secretary of Defence who championed the development of weapon autonomy in the United States armed forces, once uneasily put it, at the end of this race for speed lies the inevitable delegation of the kill decision to machines, which is undesirable from the point of view of the United States as well. China uses the term "battlefield singularity" for that moment – the point in time after which the increasing speed of battle will have outpaced human cognition and control for good.

The two major strategic risks associated with this are intensified escalation dynamics and instability.

The problem of escalation is the result of an unpredictable interaction between two or more autonomous systems. In high-frequency trading on the financial markets, unexpected interaction processes between two or more autonomously operating trading algorithms are already common, which frequently results in "flash crashes" and, accordingly, financial damage. This phenomenon can be regulated on the financial markets, but without a regulation of AWS on the battlefield that is verifiable and binding under the international law of war, a confrontation between opposing AWS could bring about risks such as unintended interaction effects, which could go as far as the unintentional use of weapons. Due to autonomous attacks and counterattacks, a "flash war" could escalate rapidly before humans could intervene with corrective measures.

Autonomy in weapon systems also promotes instability and can potentially have particularly serious consequences at the nuclear level. Entanglement is a key word in the debate about effects that arise for the strategic level, such as non-nuclear threats to nuclear weapons and their C3I systems due to the increasing capabilities of conventional weapon systems, including autonomous ones. In the area of maritime warfare, autonomy opens new possibilities for engaging enemy submarines. For example, the DARPA-funded Anti-Submarine Warfare Continuous Trail Unmanned Vessel (ACTUV) programme is currently testing the Sea Hunter, an autonomous trimaran. Its ability to detect and pursue submerged ballistic missile submarines limits the second-strike capabilities of other nuclear powers. The entanglement problem is exacerbated by an increasing willingness to use nuclear means to retaliate against non-nuclear attacks – which include, as shown above, early-warning and control systems in addition to the weapons themselves. The Trump

6 In addition, the research that is currently conducted at the interface between machine learning and computer security under the term "adversarial examples" suggests that automated image recognition offers adversaries new potential targets, for example by feeding autonomous systems manipulated sensory input or, in the case of systems that learn while in use, even "re-training" them by repeatedly deceiving them in the field.

administration's nuclear posture review signals that the United States may, from now on, respond with nuclear means to significant, non-nuclear strategic attacks. This position has been held by Russia for some time, due to the US advantage in matters of conventional weapons technology. Now, it is mirrored by the United States, likely to further adverse effects on the stability between the two largest nuclear powers.

Germany must thus have an interest that goes beyond its own armed forces in reducing the risks associated with autonomous weapon systems. This is not least because the dual-use technology on which AWS are based is vulnerable to proliferation that can extend even to non-state actors.

Recommendations: Exploit advantages, avoid risks

As signalled by the current coalition agreement, the German government is opposed to autonomous weapon systems that are not controlled by humans and continues to promote a worldwide ban on these systems – a position that was stressed again by Lieutenant General Ludwig Leinhos at this year's Munich Security Conference. On the one hand, autonomy in weapon systems does offer a number of military advantages. On the other hand, however, it involves operational and strategic risks if the targeting cycle – including the selection and engagement of targets – is completely beyond human control. A nuanced, well-considered use of autonomy in weapon systems is therefore advisable, with particular care given to the last two phases of the targeting cycle.[7]

This study therefore recommends seizing the advantages of autonomy in the non-critical phases of the targeting cycle (such as navigation, target recognition, etc.). At the same time, the risks of an autonomous selection and engagement of targets should be prevented both on the level of doctrine and of arms control policies. Recommended measures are:

Measures with regard to operational risks

- Develop and publish a policy document (similar to what allies such as the United States or the United Kingdom have done) that is initially valid for five years before it becomes subject to review, in which the Bundeswehr determines how to use autonomy in weapon systems. This document should prescribe that the last two phases of the targeting cycle always remain under meaningful human control; in other words, the Bundeswehr should not field any weapon systems that are, according to this definition, "fully autonomous".

- Define and regulate the continued use of precisely defined defensive systems[8] as an exception to this rule.[9]

- Initiate a research program to determine the parameters necessary to maintain human control of weapon systems in future human-machine-interactions in weapon systems (e. g. the Future Combat Air System). Research question: To what extent can human decision-making and control as part of the targeting cycle be supported by machines before it loses its essence?[10]

Measures with regard to strategic risks

- Identify and establish confidence-building measures and best practices to maintain human control of weapon systems in cooperation with friendly and allied armed forces.

- To prohibit (aside from in defence systems) the selection and engagement of targets without human control, continue and intensify German efforts to develop an international system of rules binding under international law for autonomy in weapon systems.

- Explore possibilities for effective arms control using methods suited to verifying such a prohibition.

7 The implications of autonomy in earlier phases of the targeting cycle could be the subject of a later Metis study. This should include a more detailed analysis of how to design a division of labour between human and machine for these phases, one that strikes a balance between legal, ethical and military considerations.

8 For example, "stationary systems that are limited to the automated execution of a small and precisely predefined number of pre-programmed actions and that may be used in simple environments solely under extreme time pressure in response to incoming fire and against inanimate military objects".

9 Possible additional exceptions for legacy systems meeting the functional definition of AWS, such as specific types of naval mines, for example, would have to depend on a case-by-case review.

10 See footnote 7.

Metis

IMPRINT

Publisher

Pilot Project Metis
Bundeswehr University Munich
go.unibw.de/metis

Author

Dr. Frank Sauer
metis@unibw.de

Art Director

Christoph Ph. Nick, M.A.
c-studios.net

Photography

Arif Wahid on Unsplash

Original title

Sicherheitspolitische Auswirkungen der Digitalisierung: Zukünftige Konfliktformen und Konfliktbearbeitung

Translation

Federal Office of Languages

ISSN-2627-0609