The Hague
Centre for
Strategic
Studies

SECURITY PROGRAM

Capstone Report

Robotic and Autonomous Systems in a Military Context

The Hague
Centre for
Strategic
Studies

HCSS helps governments, non-governmental organizations and the private sector to understand the fast-changing environment and seeks to anticipate the challenges of the future with practical policy solutions and advice.

Capstone Report
Robotic and Autonomous Systems in a Military Context

Security Program
The Hague Centre for Strategic Studies

ISBN/EAN: 9789492102850

Authors: Michel Rademaker, Amit Arkhipov-Goyal, Salma Atalla, Frank Bekkers, Patrick Bolder, Esther Chavannes, Alen Hristov, Hugo Klijn, Klaudia Klonowska, Maaike Okana-Heijmans, Michel Roelen, Tim Sweijs and Bianca Torossian

January 2021

Design: Mihai Eduard Coliban (layout) and Constantin Nimigean (typesetting).

The Hague Centre for Strategic Studies
info@hcss.nl
hcss.nl

Lange Voorhout 1
2514EA
The Hague
The Netherlands

SECURITY PROGRAM

Capstone Report

Robotic and Autonomous Systems in a Military Context

HCSS Project Team

Michel Rademaker
Deputy Director and Project Leader

Frank Bekkers
Director of Security

Tim Sweijs
Director of Research

Bianca Torossian
Strategic Analyst

Lt. Kol.
Patrick Bolder
Strategic Analyst

Esther Chavannes
Strategic Analyst

Lt. Kol.
Michel Roelen
Strategic Analyst

Amit
Arkhipov-Goyal
Junior Strategic Analyst

Klaudia Klonowska
Junior Strategic Analyst

Sounding Board

Casper Chorus (TUD)
Pieter Elands (TNO)
Martijn Hädicke (RNLA)
Sandra de Jongh (MinBuza)
Daan Kayser (PAX)
Hugo Klijn (Clingendael Institute)
Kees Matthijsen (RNLA)
Sjoerd Mevissen (RNLA)
Norbert Moerkens (RNLA)
Miriam Otto (MinBuZa)
Martijn Reubzaet (MinBuZa)
Lukas Roffel (Thales)
Egbert-Jan Sol (TNO)
Miriam Struyk (PAX)
Ilse Verdiesen (RNLA, TUD)
Arthur van der Wees (Arthur's Legal)
Wim Zwijnenburg (PAX)

Authors

Michel Rademaker, Amit Arkhipov-Goyal, Salma Atalla, Frank Bekkers, Patrick Bolder, Esther Chavannes, Alen Hristov, Hugo Klijn, Klaudia Klonowska, Maaike Okana-Heijmans, Michel Roelen, Tim Sweijs and Bianca Torossian

Factsheets by Jelle van der Weerd
Synthesis by Bianca Torossian and Michel Rademaker

Chapters

About the RAS Project

Militaries around the world are developing and using robotic and autonomous systems (RAS) and the conditions under which this process takes place within the Netherlands and what opportunities and challenges are likely to arise as a consequence is of great importance. The HCSS project 'RAS in a Military Context' sought to contribute to this discussion. Over a two-year period, the project yielded five public research papers covering a range of topics relevant to the implementation of RAS in a military context. These research papers cover military applicability, ethical considerations, legal discourse, requirements for cooperation and the implementation of RAS in a military context. All five papers are combined in this Capstone document, including a Synthesis, which briefly summarizes the analyses, and a series of six factsheets.

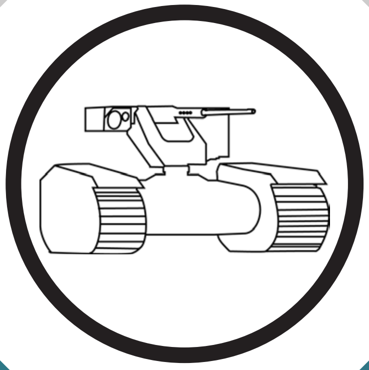
Our approach was focused on acquiring the expertise of practitioners, researchers, ethicists, legal specialists, industry professionals, technicians, civil society organizations, military personnel and other members of the defense community. Doing so enabled us, not only to gather a multi-faceted understanding of the subject matter, but also to uniquely connect these stakeholders together and foster challenging discussions between them. Over the course of the project we held five meetings with a diverse group of Sounding Board members who steered our research trajectory and provided valuable input into our position papers and draft research papers. We also gathered expertise from over 200 stakeholders who joined our six expert sessions, which involved various methodologies including scenario based discussions, design sessions, serious gaming exercises and interviews.

Our secondary objective was to inform public debate and create a more nuanced conversation about RAS in a military context that resisted prevailing ideas of 'killer robots'. To this end, we held public symposiums on the ethical dilemmas RAS pose, released five podcasts, organised conferences and roundtables and in February 2021, we released an 18-minute documentary, available on YouTube.

The RAS Project Team would like to thank all who have been involved in this project for generously offering their time and expertise, particularly our Sounding Board members. Our hope is that readers find our research and practical recommendations helpful in structuring their thinking and that the discussion on this important topic will continue to develop and thrive.

Michel Rademaker, Project Leader

Table of Contents



Summaries

Robotic and Autonomous Systems in a Military Context

Introduction

In December 2019 during a speech at a Russian Defense Ministry board meeting, Putin stated that "robotic systems and unmanned aerial vehicles are being rigorously introduced and used in combat training, which dramatically boosts the capabilities of armed units and subunits." A few months later, the Russian Defense Ministry announced a closed tender worth approximately 4.2 million euros that sought "Research on creating an experimental model of neural network development, training, and implementation for the new generation of artificial intelligence military systems". While China is far less boastful publicly, their strategy for military superiority is lead by evolutions in AI and automation, causing some analysts to stipulate that the PLA aims to dominate through system-of-systems conflict and highly intelligentized warfare.

These blips signify a larger phenomenon. Militaries around the world are developing, integrating and using robotic and autonomous systems in line with the forth evolution of warfare and further thinking needs to be done regarding the conditions under which this process takes place within the Netherlands and what challenges and implications are likely to arise as a consequence.

The HCSS project 'RAS in a Military Context' sought to contribute to this discussion. Over a two-year period, the project yielded five public research papers covering a range of topics relevant to the implementation of RAS in a military context. This synthesis ties these topics together and presents the most pertinent findings of the project. Observations from HCSS research on ethical requirements, legal discourse, partner cooperation, implementation and concept development and experimentation are summarized below, preceded by a primer section on the military applicability of RAS.

The Military Applicability of RAS

Robotic and Autonomous Systems (RAS) present numerous, significant and far-reaching opportunities within a military context. In order to observe the ways in which these systems are applicable in this context and evaluate their utility, some definitions and concepts need to be addressed:

Autonomy: The level of independence that humans grant a system to execute a given task. It is the condition or quality of being self-governing to achieve an assigned task based on the system's own situational awareness (integrated sensing, perceiving, analyzing), planning, and decision making. Autonomy refers to a spectrum of automation in which independent decision making can be tailored for a specific mission, level of risk, and degree of human-machine teaming. Levels of autonomy can range from remotely controlled (non-autonomous), Operator Assistance, Partial Automation, Conditional Automation, High Automation, or Full Automation.

Robot: A powered machine capable of executing a set of actions by direct human control, computer control, or both. It is composed minimally of a platform, software, and a power source.

Robotic and Autonomous Systems (RAS): RAS is an accepted term in academia and the science and technology (S&T) community and highlights the physical (robotic) and cognitive (autonomous) aspects of these systems. RAS is a framework to describe systems with both a robotic element and an autonomous element. It is important to note that each of the consecutive parts of RAS covers a broad spectrum. The 'systems'-part refers to a wide variety of physical systems over a wide range of (in our case: military) application areas. Automated software systems running on computers or networks, including 'bots', pieces of software that can execute commands with no human intervention, do not qualify as RAS because they lack a physical component. The 'robotic' part, which refers to the physical layout of the system, holds that the system is unmanned or uninhabited. All other physical aspects (size, form, whether it flies, floats or rolls, etc.) are left open.

Lethal Autonomous Weapon System (LAWS): A weapon that, without human intervention, selects and engages targets matching certain predefined criteria, following a human decision to deploy the weapon on the understanding that an attack, once launched, cannot be stopped by human intervention.

Meaningful human control (MHC): MHC encompasses (at least) the following three elements: (1) People make informed, conscious decisions concerning the use of weapons; (2) People are adequately informed in order to ensure that the use of force conforms to international law, within the scope of the knowledge that they have on the goal, the weapon, and the context in which the weapon is put to use; (3) The weapon in question has been designed and tested in a realistic operational setting and the people involved have received adequate training, in order to use the weapon in a responsible manner. MHC is a complex concept and, in many cases, the above description is not conclusive. The official Dutch standpoint is that "all weapons, including autonomous weapons, must remain under meaningful human control."

The rhetoric of "killer robots" has narrowed the public's view of robotic and autonomous systems in a military context to being exclusively about lethal use of force by highly or fully autonomous systems. In reality, RAS can be applied to numerous military functions and tasks, with various levels of autonomy in each function (See Figure 1). The broad military applicability of robotic and autonomous systems yields numerous and vast opportunities. The challenge for the years ahead is to make the most of these opportunities and wield the potential for military advantage whilst simultaneously mitigating the risks posed.

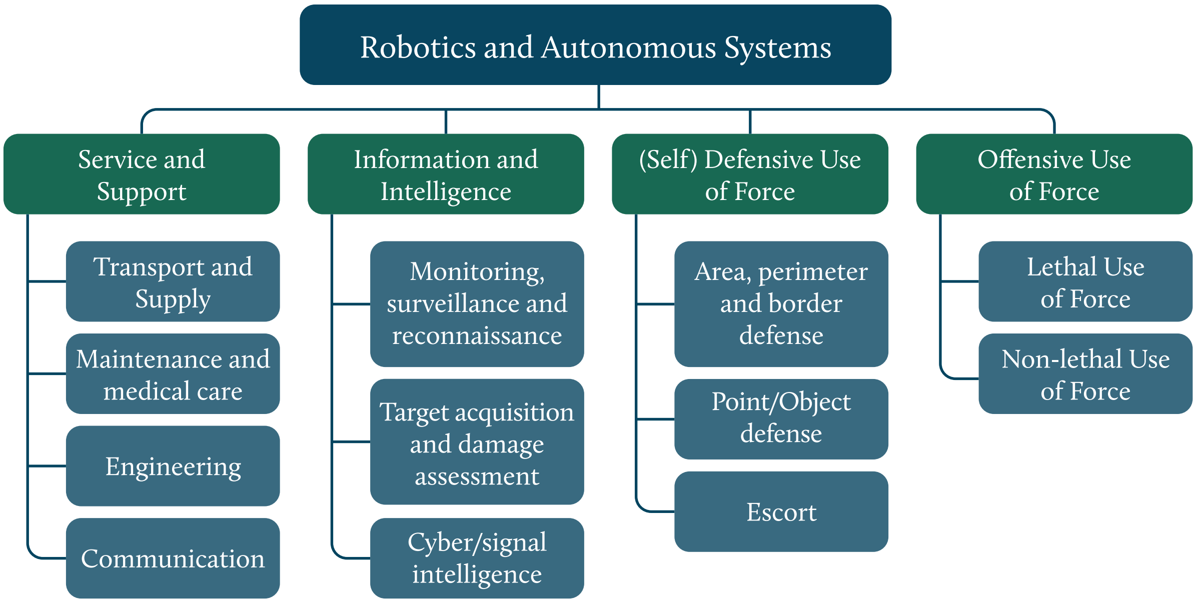

Robotics and Autonomous Systems
Service and Support
Transport and Supply
Maintenance and medical care
Engineering
Communication
Information and Intelligence
Monitoring, surveillance and reconnaissance
Target acquisition and damage assessment
Cyber/signal intelligence
(Self) Defensive Use of Force
Area, perimeter and border defense
Point/Object defense
Escort
Offensive Use of Force
Lethal Use of Force
Non-lethal Use of Force

Figure 1. The range of application areas for RAS in a military context

Implementing RAS into these functions brings significant challenges, but also heralds new opportunities for militaries to be more effective, efficient and agile. The potential of RAS to continue to (r)evolutionise the defense arena can be evaluated according to these categories.

Speed. With the help of artificial intelligence, which stimulates rapid decision-making and prioritization of threats, RAS are already capable of surpassing human reaction times and shortening the OODA (Observe, Orient, Decide, Act) loop.

Reliability. Delegating tasks to machines requires an immense degree of trust and as of yet RAS cannot prove adequate reliability across all military application areas. However, our confidence in these systems will increase as they prove their reliability and effectiveness in executing specific tasks.

Accuracy. AI systems have developed facial image recognition and sensory abilities past the level of human performance, though the claim that unmanned systems are more precise than human operatives is widely disputed.

Mass. Owing to increased range and endurance, RAS has the capability to enhance coverage of the battlespace and overwhelm adversaries. The best example of this potential is 'swarming'.

Reach. RAS greatly enhance the available points of presence for surveillance, intelligence, reconnaissance and weapons systems.

Robustness. In the short term, RAS will be more vulnerable than humans to fail due to unanticipated conditions including poor weather and changes to the mission. This frailty extends to the virtual domain: as losses in connection, hacking and other interference can render a system incapable.

Safety. RAS can perform 'dull, dangerous and dirty' tasks so that humans can focus on the more specialized tasks and be kept out of the line of fire.

Cost. Although exclusive access to the most cutting-edge technology will be reserved for the wealthiest players, the cost of systems that are now considered highly-advanced will fall throughout the next twenty years, thus becoming more widely attainable.

Maintenance. Updating and upgrading RAS software and hardware may prove more difficult given the complexity of the systems and the multiple (external) partners involved.

Time efficiency. RAS can perform dull and repetitive monitoring tasks at a high standard 24/7 without the need for rest, logistical planning can be solved efficiently, and the limits of human multitasking can be quickly surpassed.

Flexibility. Although RAS currently excel in executing specific tasks and humans will remain the most flexible for the foreseeable future. This dynamic is likely to change as developers continue to innovate current systems.

Adaptiveness. RAS are highly adaptive and be easily reconfigured during the system's life cycle (scaled, extended, upgraded etc.) over time so to keep up with new requirements emerging in a dynamic environment.

External legitimacy. The military's engagement with RAS must thus strike a balance between the advanced capabilities they (potentially) provide and the values and norms of the society it serves.

Internal legitimacy. Trust and organizational normalization of RAS will be strengthened over time. As understanding of the systems, their predictability and their familiarity grow, their legitimacy within the organisation will solidify.

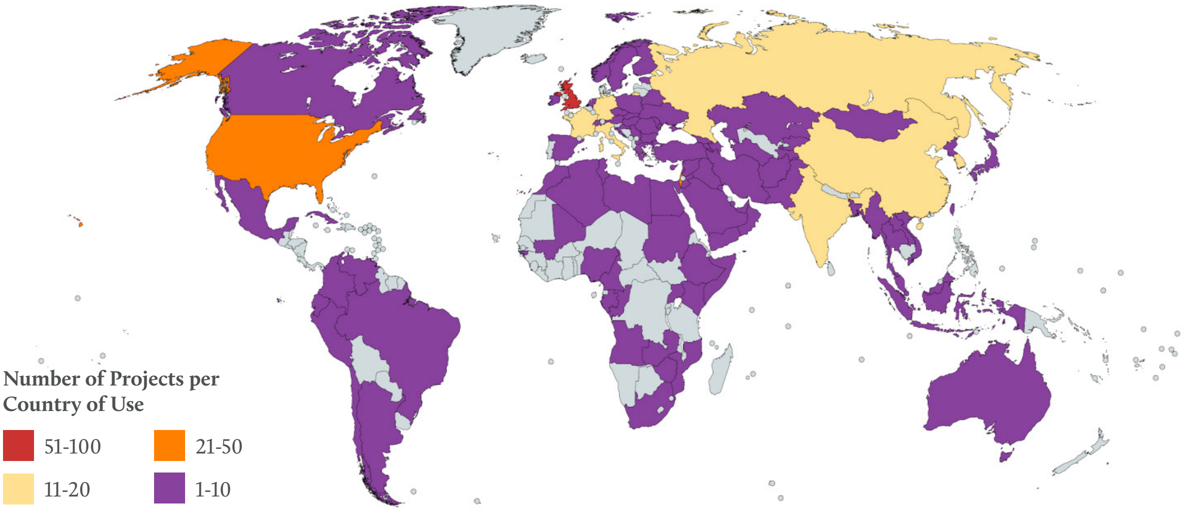
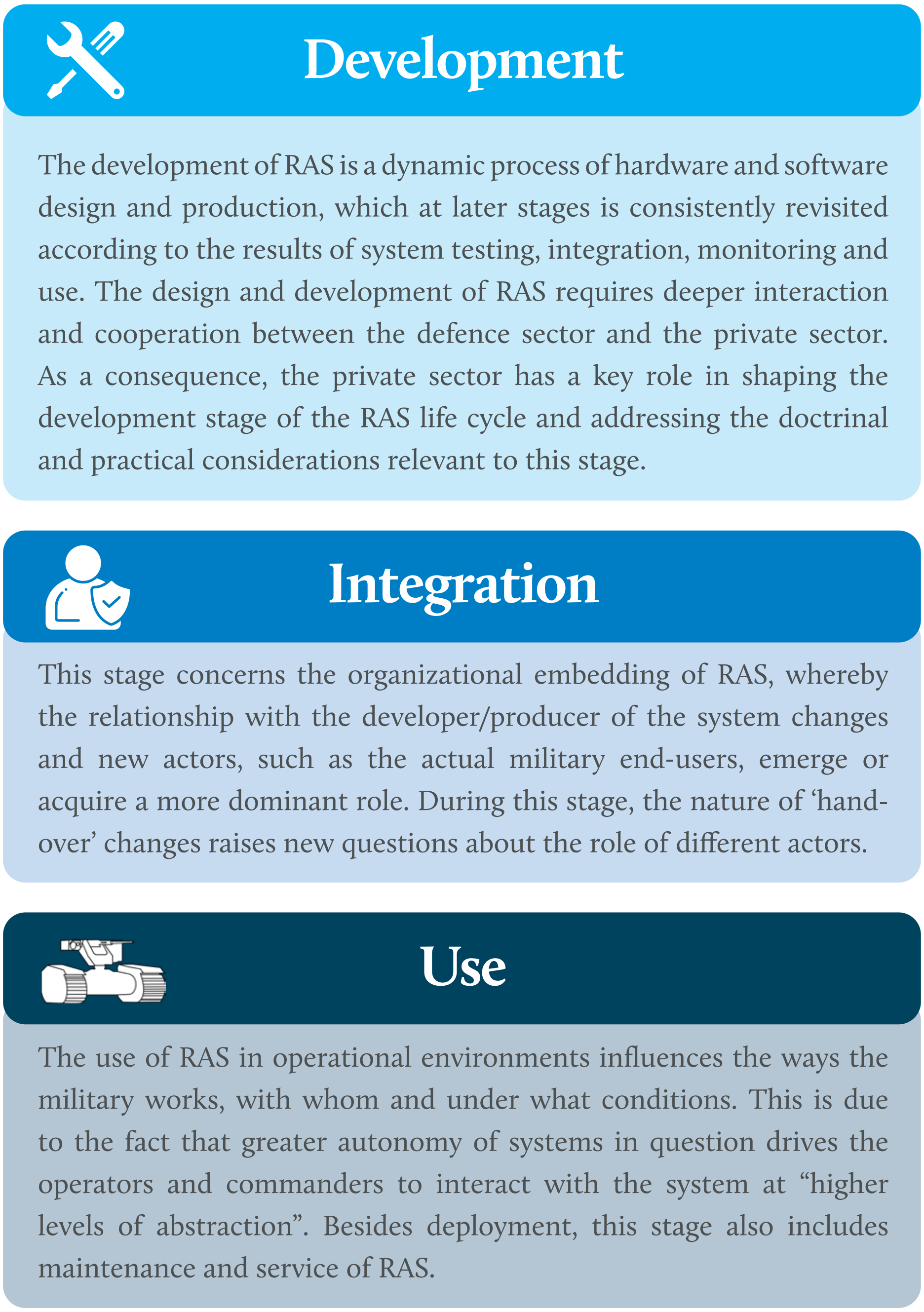


Figure 2. Number of RAS projects per country of use (2018)

Recognising this potential and sensing a need to be militarily competitive in a rapidly changing international arena, a number of states use RAS as part of their armed forces (Figure 2). However, despite the apparent opportunities, implementing RAS in a military context is no simple task. Numerous practical and doctrinal challenges mar the implementation process, demanding discussion among policymakers, innovators, researchers, the defence community, and members of civil society, and in many cases, these challenges test the very systems we depend on to regulate, develop, acquire, integrate and use other military technologies. In order to analyze these challenges, it is important to identify three stages of the RAS system life cycle: development, integration and use (Figure 3). The next section summarises the doctrinal considerations (ethical and legal) and the practical considerations (private sector cooperation and concept experimentation and development) explored and analysed during the HCSS RAS Project.[1]



Development

The development of RAS is a dynamic process of hardware and software design and production, which at later stages is consistently revisited according to the results of system testing, integration, monitoring and use. The design and development of RAS requires deeper interaction and cooperation between the defence sector and the private sector. As a consequence, the private sector has a key role in shaping the development stage of the RAS life cycle and addressing the doctrinal and practical considerations relevant to this stage.

Integration

This stage concerns the organizational embedding of RAS, whereby the relationship with the developer/producer of the system changes and new actors, such as the actual military end-users, emerge or acquire a more dominant role. During this stage, the nature of 'hand-over' changes raises new questions about the role of different actors.

Use

The use of RAS in operational environments influences the ways the military works, with whom and under what conditions. This is due to the fact that greater autonomy of systems in question drives the operators and commanders to interact with the system at "higher levels of abstraction". Besides deployment, this stage also includes maintenance and service of RAS.

Figure 3. Explanation of the RAS life-cycle. [2]

[1] This paper serves as a synthesis for number of research papers produced by HCSS on the topic of RAS in a Military Context. The aforementioned research papers delve into each topic in great detail and should be consulted for more detailed information.

[2] The specific nature of RAS often results in a spiral development process, whereby stages of the lifecycle reoccur and/or occur simultaneously. Though recognizing the complexity and interrelated nature of this process, the somewhat linear division of the RAS lifecycle into Development, Integration and Use is employed in this paper for the sake of simplicity.

Societal considerations for the military applicability of RAS

Ethical and legal considerations on the development, integration and use of RAS for a military context abound. While the current ethical debate on robotic and autonomous systems (RAS) is often dominated by relatively extreme narratives surrounding a total ban on 'killer robots', current discussions on RAS have sidelined nuances that have critical implications for deciding how to introduce RAS in a military context. The brewing AI arms race and the diffusion of cheap, technologically advanced systems among state and non-state actors compels countries to adopt RAS. How militaries can do so whilst also keeping in-tact human agency, human dignity and responsibility is of great importance.

Ethical considerations

Maintaining human agency, particularly in the context of AWS, is one of the most contentious issues of debate with respect to the integration of RAS in the military domain. Human agency is a concept that encompasses "self-control, morality, memory, emotion recognition, planning, communication and thought."[3] It includes "features of self-awareness, self-consciousness and self-authorship," and as a result relates to moral agency and affects the attribution of responsibility.[4]

Human control, also referred to as 'meaningful human control' (MHC), is an operational component of human agency, which distinguishes between human and artificial decision-making processes.[5] A fundamental aspect of maintaining MHC is the operator's understanding of the algorithmic process' parameters, the outcomes presented as a result of the computation, and the ability to explain the machine's path to conclusion after the fact. From this point of departure stems an important ethical concern of RAS, and AI in particular: the lack of algorithmic transparency. Algorithms such as neural networks suffer from opacity as they operate as 'black boxes', whereby the path taken by the algorithm to arrive at the conclusion is often not traceable.[6] The diminished understanding an operator has of such systems reduces their ability to predict and/or explain the

[3] Gray, Gray, and Wegner, "Dimensions of Mind Perception."

[4] European Group on Ethics in Science and New Technologies (EGE), Statement on Artificial Intelligence, Robotics and "autonomous" Systems.

[5] MHC interrelates with "effective control", a prerequisite in public international law for legal liability and unlawful conduct. In the context of the use of RAS/AWS, the term is used alongside "effective command" to determine state responsibility.

[6] Preece, "Asking 'Why' in AI: Explainability of Intelligent Systems – Perspectives and Challenges"; Matthias, "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata," 178–79.

system's reasoning process, undermining the control that the operator has over the outcomes and hence, the responsibility for its (mis)use. Furthermore, the evolutionary nature of algorithm-driven systems, both as a result of self-learning properties and software updates, has the potential to considerably affect explainability of systems' actions. Self-learning AI that independently develops its understanding of the surrounding environment, automation bias and excessive trust in system outputs may limit human control over a RAS system's operation. As the design of a system can incorporate various degrees of autonomy (from remote controlled to fully autonomous) within the multiple functions of a system across the Observe, Orient, Decide, Act (OODA) loop, meaningful human control principles should be considered at the earliest stages of development.

The fundamental guiding principle is to work with 'ethics by design', whereby ethical considerations are incorporated in the use case identification, system design, validation, manufacturing, and testing processes, rather than solely in the 'use' stage of the system life cycle. This entails building an understanding of the system performance and behavior early in the design and testing stages by involving end-users early, meaning that operators, supervisors and commanders will be better able to trace, understand and predict the system's decision-making process. Best practice guidelines should be created for the outsourcing of the development process to external contractors.

Legal considerations

The lack of meaningful human control in RAS results in considerations for legal discourse too and debate on governing autonomous weapons is gaining momentum. International positions still differ widely, ranging from proponents and opponents of a ban on such weapons to a group of countries that lie in between and emphasize the need for further clarification and elaboration of existing regimes. It is clear however that current rules, standards, and practices are relevant but, most probably, insufficient to cover developments with regard to autonomous weapons. At the very least, RAS would require refinements of existing regulation. While the consensus-based CCW/GGE[7] still counts as a necessary tool to further this debate, it is doubtful whether this effort alone is sufficient. Despite the inclusion of NGOs and academia, state parties are dominant in this format, and industry is only present in a backbench capacity. Amid various approaches to definitions, norms, and standards at the international level, the Netherlands need to decide on the directions of the modernization of their armed forces and their international posture amid an intensifying public debate.

Legal approaches to regulating RAS include hard law, soft law, and voluntary measures. Hard law concerns binding treaties that are negotiated and agreed upon between states. Soft law involves quasi-legal instruments such as politically binding Codes of Conduct (CoCs) or Confidence and Security Building Measures (CSBMs), sometimes involving multiple stakeholders other than states.

[7] The Group of Governmental Experts (GGE) of the High Contracting Parties to the Convention on Certain Conventional Weapons.

Finally, voluntary instruments include behavioural principles or norms and exchanges of best practices or other information, within or outside traditional arms control communities (Figure X).

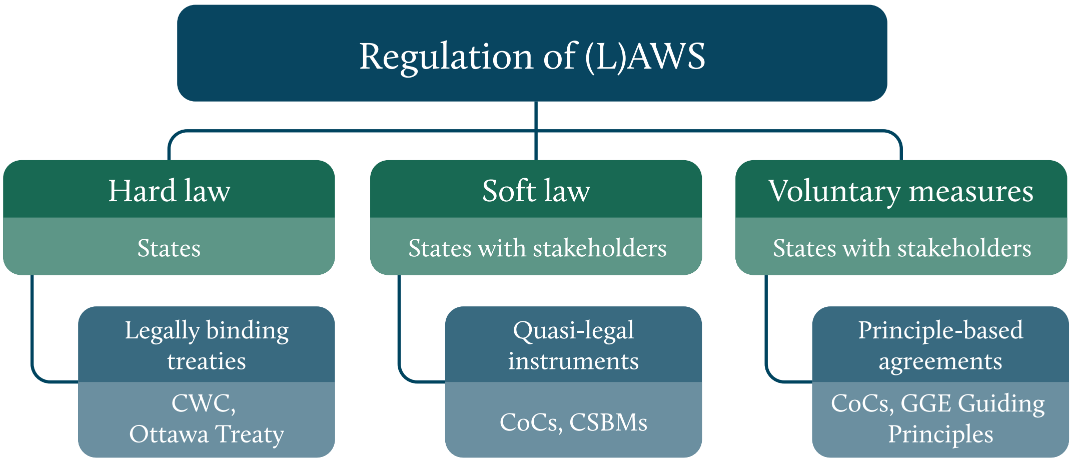


Figure 4. The three approaches to the regulation of RAS.

Numerous characteristics inherent to RAS make regulation (especially hard law approaches) particularly challenging. First, these technologies are developing fast and offer sometimes spectacular prospects for their use in both military and civilian applications, hence the temptation of some politicians and NGOs to hone in on alarmist scenarios and frame the discussion exclusively in terms of 'killer robots'. This oversimplification hinders nuanced legal debate. The fact that autonomy is not a static function of weaponry further complicates the matter, because this elusiveness means that the discussion is not always about identifiable (weapons) systems, as is the case with existing regimes that mostly regard specific categories such as chemical, biological, or nuclear weapons, or certain types of conventional arms or delivery systems. Instead, regulation needs to grapple with algorithms that are dual-use and that may or may not be harmful when implemented into a system (for example, harm may depend on an action that the system 'learns' after years of use).

Second, given the early and complex phase of the RAS debate, and the lack of common language on definitions and categorizations, hard law could be out of reach. Existing arms control regimes relate to established, well-defined weapon categories. Furthermore, the current geopolitical climate seems hardly conducive to new multilateral arms control agreements.

Lastly, technological innovation has traditionally emanated from the military-industrial complex, with innovations finding civilian applications later on (known as the spin-off effect). In the case of RAS, the trend appears to be going in the reverse direction (spin-in), and it is claimed civilian innovation facilitates the design of LAWS. In the framework of designing control regimes, 'spin-in' requires strong interaction with the private sector and will lead to forms of 'shared responsibility and accountability', which are not entirely new but will be more difficult to manage.

For these reasons, soft and/or voluntary instruments appear to be the more realistic way forward, as these are easier to reach, have lower thresholds for entry, and enable the inclusion of non-governmental stakeholders. Soft arrangements are less static and, by definition, more flexible and adaptable to new circumstances. Under existing circumstances, this type of arrangement is probably the highest attainable goal. Taking into account the forward-looking nature of this debate and the many unknowns in this respect, there is a specific focus on generating more transparency and opting for positively framed recommendations instead of more classical prohibitive measures. Trusted communities can be helpful in enhancing the debate as these networks have the ability to bring together key actors to provide input for developing principles and norms for further regulation or export control regimes that are based on mutual trust and respect.

Operational considerations for the military applicability of RAS

Operational considerations for the military applicability of RAS involve the challenges posed by existing processes and culture within the Armed forces, specifically in regard to cooperation with external partners and concept development and experimentation (CD&E). In terms of cooperation, the emergence of RAS challenges the effectiveness of multi-stakeholder cooperation in a military context, especially when it concerns collaborative structures with the private sector. In addition to changes with external relations, the Armed Forces must also grapple with internal restructuring of CD&E processes, which will call into question not only the structural processes that guide the function of the organization, but also broader doctrinal thinking. RAS are unique in that they ultimately can take humans 'out of the loop' and, as a consequence, drastically affect operational performance, organizational embedding (e.g. influencing numbers, skills and training of personnel), and operational concepts (doctrine and tactics).

Considerations for Cooperation

As RAS development is to a great extent driven by civilian innovation, the integration of RAS creates demands for interaction with designers, developers, and manufacturers outside the traditional defense industry. Managing this relationship requires a) integrated and interdisciplinary cooperation, b) a clear division of tasks, investments and responsibilities, c) the implementation of common system architecture and d) a balancing of military requirements and expectations, technological possibilities, and (potentially conflicting) legal, ethical and safety parameters.

Additionally, due to a rapid cycle of innovation within, for example, artificial intelligence (AI), RAS must be developed and acquired in fast-paced procedures, used for shorter periods of time, and modified, updated, inserted, or exchanged throughout the life cycle of the system. Whereas the integration of a regular system includes some sort of 'hand-over' from the developer/producer to the military organization who will use the new system, a feature of RAS is the dependency upon integrated software that continues to evolve; certainly, where self-learning algorithms are part of the autonomous reasoning of the system. As a result, the hand-over of RAS does not necessarily finalize the involvement of the producer in the latter stages of the life cycle. The producer must ensure that the system is adequately and regularly updated, and that the self-learning nature of the system is controlled and continues to meet demands and standards.

In contrast to these rapid cycles of innovation, societal discussions regarding ethical questions and legal uncertainties unfold slowly. These conversations demand interaction with a range of stakeholders and policy makers external to defense organizations. Militaries should seek to involve external developers with this debate as much as possible and should exercise meaningful oversight over all stakeholders through the use of internal guidelines or codes of conduct.

The integration of RAS involves the adaptation of all the 'DOTMLPF' categories.[8] The military should reconsider whether the doctrine covers the situations of RAS deployment, whether the training and organization of forces are sufficient to ensure that RAS is taken full advantage of, whether there is sufficient technical literacy to deal with ad-hoc technical problems, whether the facilities are equipped to repair RAS, etc. The fundamental changes that RAS might bring to (some or all of) the DOTMLPF-elements requires broad interaction with stakeholders within the defense organization, with international military partners, and possibly with other partner agencies.

Considerations for Concept Development and Experimentation

The introduction of RAS within the Armed Forces constitutes more than just getting used to and working with new weapon systems. To be at the forefront of quickly changing needs and emerging technologies and to be able to make the right decisions on how RAS enhance the Army, experimentation is key. For developers of the materiel, the military world can be rather new and as new issues are arising when working with RAS, intensive working relations with developers, producers, knowledge institutions and of course the operational users themselves have to be organized. These working relations and the discussions that follow should be established for products that are almost ready, but also especially for the Armed Force's most conceptual ideas. These discussions could take place in so-called 'testbeds' ('proeftuinen').

Within the Armed Forces, the culture of how to shift from the current and planned force towards the future force requires an attitude change. In-depth conversations and research needs to take place, not only on topics of certainty, such as updating older equipment, but also on uncertainties such as thinking of capabilities that will be required in the future and strategizing on how to reach that point. The defence planning system needs to be adapted for this and may also require a separate innovation fund within the Defence Investment or Lifecycle Plan. First and foremost, the strict rules on procurement will need some tweaking in order to allow for an innovative transition from older means to new ones. One possible approach could be to form a working group at the early stages of defining new demands for capabilities, where all players concerned, from legal teams and acquisition support, to the operational user and the responsible staff officer, can converse and plan for the acquirement of new capabilities.

[8] Doctrine, Organization, Training, Materiel, Leadership, Personnel and Facilities.

Our interviews found that quite a few innovations stem from the lower levels of the organisation, rather than being the product of a bigger-picture thinking on future technologies. Innovation departments should provide for flexible procedures and a certain level of freedom to manoeuvre in order to allow for 'grassroots' innovation. Innovation competitions are another good example of how the organisation can rather quickly and without complex procedures reach certain innovative solutions. Also of importance is the idea that innovations are allowed to fail and that in such a case, nobody has to be punished for such failure. Though it can not be expected for military organisations have the same acceptance of failure and appetite for risk that can be found in industry, a shift must occur toward this attitude if progress is to occur. Failure and risk are inalienable components of true innovation.

Years of decreasing budgets for the Armed Forces and ever-stricter rules on budgeting, is not a conducive backdrop for exploring new and untrodden paths such as RAS. What does help however, is the fact that the whole organisation, from high to low sees the necessity of innovation and innovative projects to deliver the most valued product of the Armed Forces: fighting power, now and in the future.

Conclusion

Robotic and Autonomous Systems represent a transformation in military domain. They provide a significant military capability to extend the quality, reach and efficiency and safety of military operations and are changing the way we fight conflict now and in the future. RAS capabilities are being implemented, not only by the Dutch Armed Forces, but also by our potential adversaries.

The technical, operational, legal and ethical issues, as well as the potential proliferation of this emerging technology are complex and relatively new. As new developments come to light and experiences are gained during operational use, the way in which we conceptualise, design, build and operate RAS will require iterative reconsideration.

This also means that the need for constant knowledge development, concept development and experimentation is required. Real-life testing in operational settings is necessary to fully understand the potential of RAS and its requirements to be utilized as an important strategic tool in the military toolbox.

For industries to fully engage in these markets, co-development and co-experimentation in this fast-developing field requires a different mindset. Until the products are fully matured and operational use shows that they are well understood and predictable, RAS will require constant adjustments through short-cycle innovation processes.

The development and implementation of RAS in a military context will require constant attention, creative big-picture thinking and strong collaboration networks with stakeholders, including policymakers, academics, ethicists, lawyers, industry professionals, technicians, civil society and the defence community.

Factsheet

The Military Applicability of Robotic and Autonomous Systems

The increasing potential of RAS

RAS presents numerous, significant and far-reaching opportunities, including creating better situational awareness, reducing the physical and cognitive loads of soldiers, sustaining the force, and extending the reach and persistence of operations. What is the current state of RAS in the military context and what is the potential?

The military value of RAS

Effectiveness – RAS allows rapid action, accuracy, has extended reach and can perform dull, dirty and dangerous tasks

Efficiency – RAS is extremely time efficient and can substantially increase cost effectiveness

Agility – RAS allows for greater flexibility, but adaptiveness is highly dependent on the system used

Legitimacy – RAS must serve the values and norms of society while also being trusted and predictable

Current RAS Applications

Contrary to popular belief, the majority of RAS are used for Information & Intelligence and Service & Support Tasks, not for Use of Force (Figure 1). Moreover, RAS is being developed globally - not only by great powers (Figure 2).

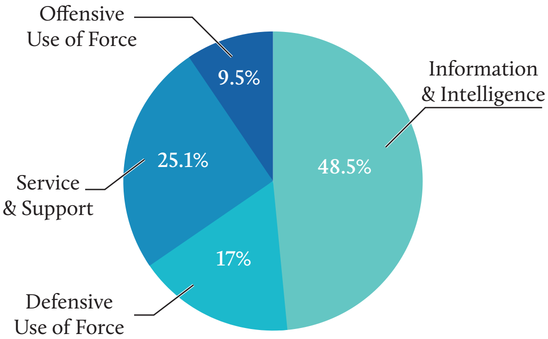


Figure 1: Applications of RAS (SIPRI: 2018)

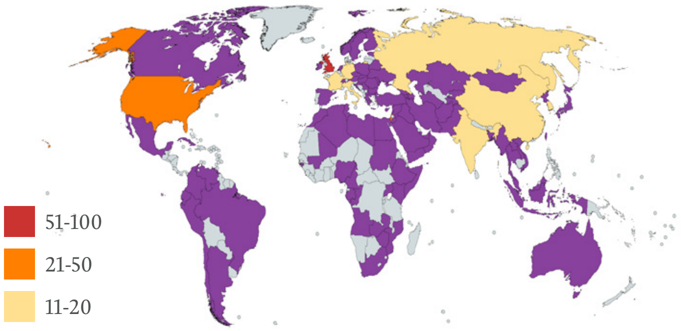


Figure 2: RAS Projects Worldwide (SIPRI: 2018)

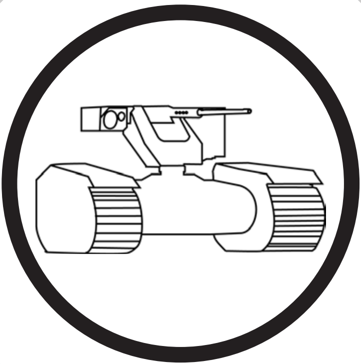
Opportunities and risks

While RAS present significant opportunities, they come with technical, personnel, doctrinal, legal,ethical, military and political challenges

Next steps: recommendations for the NLDA

- Set in motion urgently needed CD&E within the armed forces
- Ensure cyber- and electromagnetic spectrum control
- Investment into AI and systems engineering
- Organize education and training programs
- Train, improve and maximize AI with the use of large, high quality data sets
- Develop a partnership approach between private sector and military
- Adapt procurement processes toward buying smart systems
- Ensure that systems are predictable, familiar, understandable and appropriate

The Hague Centre
for Strategic Studies
Lange Voorhout 1
2514 EA The Hague
The Netherlands
hcss.nl



The Hague Centre for Strategic Studies

Chapter 1

The Military Applicability of Robotic and Autonomous Systems

Bianca Torossian, Frank Bekkers, Tim Sweijs, Michel Roelen, Alen Hristov and Salma Atalla.

1. Introduction

1.1 This document

This paper assesses the military utility of robotic and autonomous systems (RAS) and the risks and opportunities associated with the development and use of this technology in a military context. For the purposes of this paper, the time horizon is set at five to ten years into the future and the scope of the military application areas pertains to land operations performed by the Royal Netherlands Army (RNLA).[1]

This paper is organized according to the following structure. In this chapter, we offer our demarcation of RAS, which lacks an internationally accepted definition; give a classification of distinct levels of autonomy of RAS solutions in a military context; and present a taxonomy of military functions for which RAS solutions may be deployed. Chapter 2 presents a classification of RAS in a military context as the basis for a structured discussion. Chapter 3 is dedicated to the evaluation criteria that may be used to assess the military utility of RAS solutions. In Chapter 4, based upon an extensive dataset, we map current RAS and RAS under development on the taxonomy of military functions provided in Chapter 2, in order to gain a feel for the current military applicability of RAS. In this chapter, we further hypothesize which future systems will be available in the coming five to ten years. Chapter 5 outlines the main risks and opportunities associated with the development and military use of this technology from practical and societal perspectives. Chapter 6 concludes by setting out the necessary steps required for the successful implementation of RAS in the military.

1.2 Approach

The content of this paper is derived from two complementary and interacting approaches. The first approach comprises an extensive review of the relevant literature. From this literature review, our classification of RAS in a military context and our list of evaluation criteria for the military utility of RAS were derived. This overview was largely based on *The SIPRI Dataset of Autonomy in Weapons Systems*[2] with some minor adjustments.[3] The SIPRI dataset was further refined, categorized according to our taxonomy of military functions, and augmented with additional systems from other sources, resulting in a dataset comprising of 299 systems.[4]

The second approach was focused on acquiring the expertise and experience of practitioners, researchers, legal specialists, ethicists and members of the defense community. This practical approach manifested in a workshop attended by 55 experts from industry, the defense community, academic and research institutes and the wider unmanned systems community.[5] The following questions were posed to the workshop participants (working in six separate groups):

1. What is the military utility of **RAS for different military functions** (Service & Support, Information & Intelligence, Use of Force), both in themselves and/or to augment or substitute more human centric solutions? What level of autonomy is feasible or required (possibly a growth path)?
2. What are the **technical, organizational and doctrinal - i.e. practical - issues and challenges** in the actual implementation of the suggested (high-utility) RAS solutions? What are critical steps/actions to be taken (now) in order to deal with these issues and challenges?
3. What are the **ethical, legal and societal - i.e. conditional - issues and challenges** in the actual implementation of the suggested (high-utility) RAS solutions? What are critical steps/actions to be taken (now) in order to deal with these issues and challenges?

The insights of the participants were noted and analyzed, and have primarily contributed to section 4.2 (future military applications of RAS) and chapters 5 (opportunities and risks) and 6 (next steps). The workshop was also instrumental in gauging the framework in the chapters 2 and 3.

[2] Boulanin and Verbruggem, "Mapping the Development of Autonomy in Weapon Systems."

[3] Our overview also made use of systems identified in the British Army Innovation Technology Book (BAITB), as well as studies conducted by the U.S. Congressional Research Service, Zhifeng Lim, and Boulanin & Verbruggen. "Army Warfighting Experiment 2018: Autonomous Warrior"; Feickert et al., *U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI)*; Boulanin and Verbruggem, "Mapping the Development of Autonomy in Weapon Systems"; Lim, "The Rise of Robots and the Implications for Military Organizations."

[4] It is important to note the limitations that affect the process of open source data collection. It can be assumed that countries may classify information on RAS and RAS developments for national security purposes. Therefore, while the majority of RAS are categorized under 'Information and Intelligence', while 'Use of Force' has the least RAS solutions, it may be that the actual amount of RAS for Use of Force is broader than what is known due to the classified status of projects.

[5] The expert workshop was held at The Unmanned Systems (TUS) Expo in Rotterdam on the 18th of January 2019. The participants were split into five focus groups, each of which were lead by a member of the HCSS project team.

2. Categorization of RAS in a military context

2.1 Demarcation of RAS

There is no single internationally accepted definition of RAS. However, for the purposes of this paper, the following definitions describe the concepts most accurately.[6]

Autonomy: The level of independence that humans grant a system to execute a given task. It is the condition or quality of being self-governing to achieve an assigned task based on the system's own situational awareness (integrated sensing, perceiving, analyzing), planning, and decision making. Autonomy refers to a spectrum of automation in which independent decision making can be tailored for a specific mission, level of risk, and degree of human-machine teaming.

Robot: A powered machine capable of executing a set of actions by direct human control, computer control, or both. It is composed minimally of a platform, software, and a power source.

Robotic and Autonomous Systems (RAS): RAS is an accepted term in academia and the science and technology (S&T) community and highlights the physical (robotic) and cognitive (autonomous) aspects of these systems. RAS is a framework to describe systems with both a robotic element and an autonomous element.

It is important to note that each of the consecutive parts of RAS covers a broad spectrum. The 'systems'-part refers to a wide variety of physical systems over a wide range of (in our case: military) application areas. Automated software systems running on computers or networks, including 'bots', pieces of software that can execute commands with no human intervention, do not qualify as RAS because they lack a physical component. The 'robotic' part, which refers to the physical layout of the system, holds that the system is unmanned or uninhabited. All other physical aspects (size, form, whether it flies, floats or rolls, etc.) are left open. The 'autonomous' part, which refers to the cognitive design of the system, covers the full range from fully controlled by a remote human operator to fully controlled by internal logic, i.e. the 'program' or 'software' that determines the system's behavior. In a military context, it is important to distinguish the overarching category of RAS from the much smaller category of lethal autonomous weapons systems (LAWS). Only a small fraction of the full scope of military RAS involve LAWS. Again, there is no agreed definition. The following description is the 'working definition' that the Netherlands put forward in the ongoing debate on autonomous weapon systems that takes place within the Convention on Certain Conventional Weapons (CCW):

> **Lethal Autonomous Weapon System (LAWS)**: A weapon that, without human intervention, selects and engages targets matching certain predefined criteria, following a human decision to deploy the weapon on the understanding that an attack, once launched, cannot be stopped by human intervention.

2.2 Levels of autonomy

The 'autonomous' part of RAS is the most discussed and most constrained. A crucial notion is **meaningful human control** (MHC). The formal Dutch standpoint is that "all weapons, including autonomous weapons, must remain under meaningful human control." Again, there is no internationally accepted definition. MHC encompasses (at least) the following three elements:[7]

- People make informed, conscious decisions concerning the use of weapons;
- People are adequately informed in order to ensure that the use of force conforms to international law, within the scope of the knowledge that they have on the goal, the weapon, and the context in which the weapon is put to use;
- The weapon in question has been designed and tested in a realistic operational setting and the people involved have received adequate training, in order to use the weapon in a responsible manner.

Yet, MHC is a complex concept and in many cases the above description is not conclusive. Likewise, the often used distinction between human-in-the-loop, human-on-the-loop and human-out-of-the-loop does not suffice. These terms refer to the relationship between an unspecified human and an unspecified decision loop, whereas in reality a number of different humans may relate to a number of different loops. Many of these loops are non-operational, e.g. play out in the design phase of RAS. Also, these terms cover the aspect of human control (or machine freedom). Two other concepts also embedded in the term autonomy are the complexity of the machine and the type of decision being automated.

For our purposes, we propose a taxonomy (see *Table 1*) based on the SAE international standard J3016, which identifies six levels of driving automation to categorize self-driving cars.[8] We have slightly adapted that standard to fit our context of military use. The column labelled 'monitoring the environment' specifies whether a human operator must monitor the environment in which the machine performs its task in order to decide 'the next step' at crucial decision points; or to overwrite the automated logic if something goes wrong. The column labelled 'fall-back performance' indicates

[7] Horowitz and Scharre, "Meaningful Human Control in Weapon Systems," 4 This definition (in Dutch translation) is also used in the AIV/CAVV report, Autonome wapensystemen. De noodzaak van betekenisvolle menselijke controle, from October 2015.
[8] SAE International, "Automated Driving: Levels of Driving Automation Are Defined in New SAE International Standard J3016."

what happens when unexpected situations arise: does the operator or the automated system decides how to (re)act? The 'task performance modes' column indicates whether for certain functional aspects the level of autonomy can be switched back in order to increase human involvement.

Level of autonomy	Description	Execution of core task	Monitoring environment	Fall-back performance	Performance modes
0: Remotely Controlled	the full-time performance by the operator of all aspects of the dynamic core task,[9] even when enhanced by warning or intervention systems	operator	operator	operator	n/a
1: Operator Assistance	the mode-specific execution by an assistance system of certain functional aspects[10] of the core task, using information about the environment, while the operator performs all remaining aspects of the core task, and with the expectation that the operator will respond appropriately to a request to intervene	operator / system	operator	operator	some modes
2: Partial Automation	the mode-specific execution by an assistance system of all functional aspects of the core task, using information about the environment, and with the expectation that the operator will respond appropriately to a request to intervene	System	operator	operator	some modes
3: Conditional Automation	the mode-specific execution by one or more assistance systems of all functional aspects of the core task, using information about the environment, and with the expectation that the operator will respond appropriately to a request to intervene or/ and can override the autonomous behavior	System	system	operator	some modes
4: High Automation	the mode-specific execution by one or more assistance systems of all functional aspects of the core task, using information about the environment, even if the operator does not respond appropriately to a request to intervene	System	system	system	some modes
5: Full Automation	the full-time performance by an automated system of all aspects of the core task under all environmental conditions to at least the same level as can be managed by an operator	system	system	system	all modes

Table 1: Levels of Autonomy

[9] The dynamic core task of a RAS is the task performed in direct connection to the mission the RAS was set to do. For military applications, these various tasks can be derived from the categorization above. For a cargo drone, for instance, this core task would be navigating safely to a drop-of location, delivering the cargo intact, and return home. For a surveillance drone this would be to spot and track moving targets that fit certain characteristics.

[10] It is assumed that a core task can be broken down in functional aspects in a modular fashion. E.g. for the cargo drone the core task would consist of a navigation part (to reach the drop-of location; as well as return home) and a drop-of part (deliver the cargo).

A further subcategorization for levels four and five to distinguish between levels of MHC is conceivable. An example might be the presence or lack of an 'override switch' that allows a human operator to abort the automated mission. Another issue could be the extent in which the system is able explain its reasoning in deciding on a particular course of action (in advance, in real time or afterwards). This form of transparency is important for evaluation and possibly correction of the autonomous logic.

2.3 Taxonomy of military functions for RAS deployment

We propose a three tier taxonomy of military functions that may be performed using RAS-solutions, facilitating discussions at various levels of abstraction/granularity (*Figure 1*). The first level consists of four broad categories. The categories at the second level are listed alphabetically. These two levels are comprehensive, i.e. intended to cover the full range of all possible military applications of RAS that we might think of as fitting under one or more of the level 1 and level 2 categories. At the third level, the categorization is not fixed. At that level of detail, a great number of detailed military functions for RAS may arise, including possible new ones that have currently no 'manned' equivalent (because it is too dirty, dull and/or dangerous for humans to perform). The subcategories at level 3 given below are to be considered as representative examples.

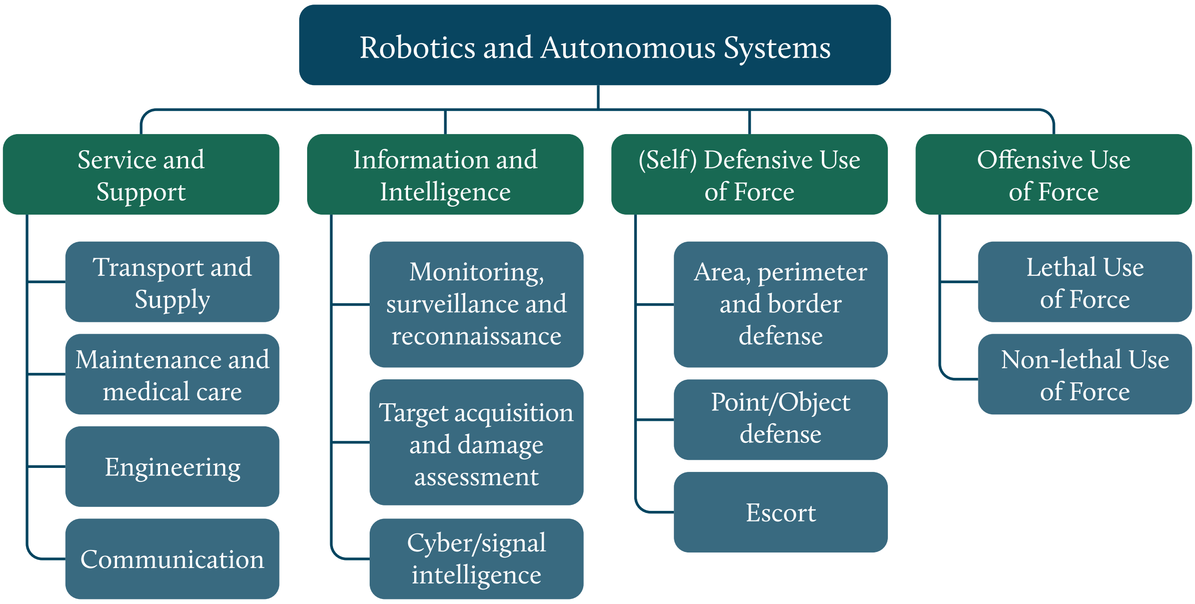

Robotics and Autonomous Systems
Service and Support
Transport and Supply
Maintenance and medical care
Engineering
Communication
Information and Intelligence
Monitoring, surveillance and reconnaissance
Target acquisition and damage assessment
Cyber/signal intelligence
(Self) Defensive Use of Force
Area, perimeter and border defense
Point/Object defense
Escort
Offensive Use of Force
Lethal Use of Force
Non-lethal Use of Force

Figure 1. HCSS Taxonomy of military functions for RAS (Level 1 & 2)

1. RAS applied in **Service & Support** activities is roughly equivalent to what the military call 'combat service support'. The execution of these activities typically resides in non-combat units. Indeed, most of these activities have non-military equivalents, with dual-use systems being feasible. For this main category, technological developments, as well as issues regarding rules and regulations and certification, are largely driven by civil sector applications. This implies that future military

RAS applications in this functional category tend to be inspired by, and make extensive use of, civil innovations. This category comprises the following level 2 functions:

a) **Transport & Supply** includes transport and salvage & recovery of materiel; transport, search and rescue and (medical) evacuation of personnel; storage and distribution of supplies (e.g. aerial refueling); trash collection and recycling; and navigation support. In geographical terms, transport & supply is to and from theatres of operation as well as within those theatres.

b) **Maintenance / Medical Care** is particularly aimed at actual operations. Outside of operations, this function typically merges with non-military (general) maintenance and medical care functions.

c) **Engineering** includes construction and demolition; mobility and counter-mobility measures; and clearance of mines, IEDs and explosives (Explosive ordnance disposal, EOD).

d) **Communication** includes all activities in support of creating or supporting one's own facilities for communication. E.g. mobile radio repeaters.

2. RAS applied in **Information & Intelligence** activities for the gathering (sensing) and processing of information in support of military planning, situation awareness & situation understanding (SA/SU) and decision making. Many of these activities have non-military equivalents, with dual-use systems feasible. However, military applications often represent a specific high-end niche, with advanced technological developments still largely set by military uses. This category comprises the following level 2 functions:

a) **Monitoring, surveillance and reconnaissance** includes observing the wider sea-land-air/space environment for potential threats, incidents, security breaches etc.

b) **Target acquisition and battle damage assessment** is distinct from the previous one, in the sense that its focuses on designated targets.

c) **Cyber/signal intelligence** pertains to information gathering and intelligence production in the electronic domain, both in the digital / cyber infrastructure and in the electromagnetic spectrum.

3. RAS applied in **(Self) Defensive Use of Force**. This category includes the use of force in response to a clear and present danger to the system itself or to a defended asset or area. This response is typically aimed at incapacitating the incoming threat, such as a missile or a projectile. This category has limited equivalents in civil security. Technological developments, as well as issues regarding rules and regulations and certification, are therefore largely driven by military applications. This category comprises the following level 2 functions:

a) **Area / perimeter / border defense** pertains to a geographically extended defense.

b) **Point / object defense** of a single object such as a building or a confined military position, as well as self-defense of the system itself.

c) **Escort** pertains to the defense of moving objects such as convoys.

4. RAS applied in **Offensive Use of Force**. This category pertains to the use of force with the explicit aim to deliberately incapacitate or kill people or deliberately damage or destroy objects (without necessarily being provoked). This category has no or little non-military equivalents. Technological developments, as well as ethical and legal issues, are therefore almost exclusively driven by military uses. This category comprises the following level 2 functions:

   a) **Lethal use of force** with the intention to kill or destroy the target.

   b) **Non-lethal or less-lethal use of force** with the explicit intention to (temporary) incapacitate the target.

3. Assessing the military value of RAS

In order to gauge the added value of RAS to the RNLA, it is necessary to identify the different ways in which these systems can (or cannot) positively contribute to the capabilities of the organization. This safeguards against innovation for innovation's sake and frames the development of RAS in terms of its potential to produce tangible, perceivable outcomes for the RNLA. In order to determine the military utility of RAS to the RNLA, we propose the following criteria (see *Figure 2*):

1. **Effectiveness** to achieve the desired effect(s) or objective(s) for the military task(s) / mission(s) where RAS is deployed.
2. **Efficiency** in the use of resources. Ideally, both the life cycle costs of the system (initial investment, maintenance, upgrades, etc.) as well as running costs (e.g. for fuel, spare parts and repair) are taken into account.
3. **Agility** to adapt according to the requirements of the situation at hand, and also to adapt over time to new situations.
4. **Legitimacy** of the application of RAS, both in a formal sense and as perceived by the (military) operators and by the people/societies, in theatre as well as at home.

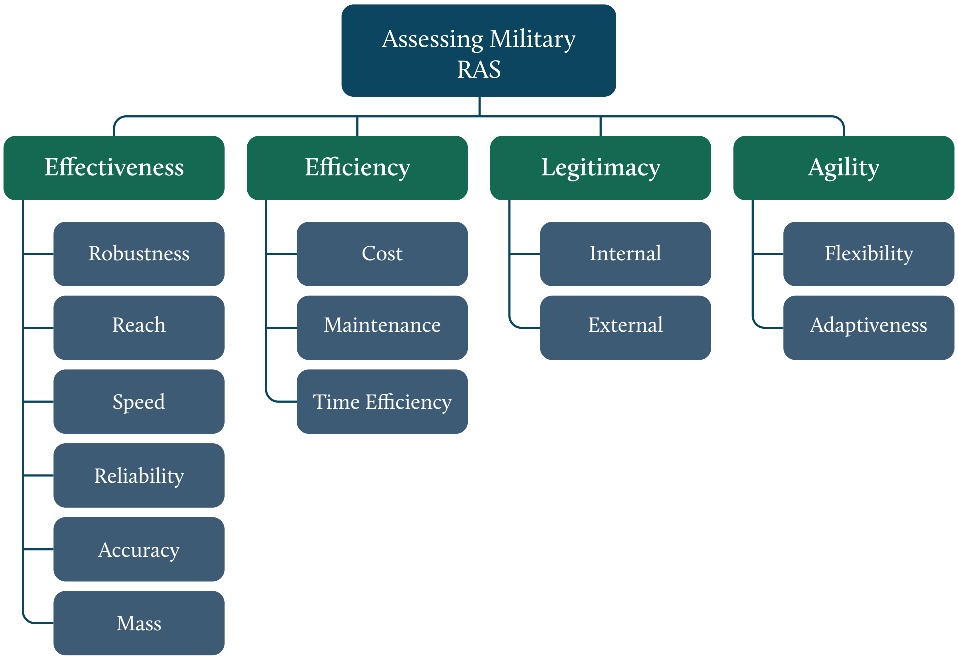


Figure 2. Evaluation Metrics used to Assess RAS

Each criterion is further broken down into sub-criteria that seek to measure whether and how RAS solutions generate added value for the military. Given the variety of military functions to consider, it is impossible to give generic absolute levels of performance for each of these criteria. What *is* possible, is for concrete applications to gauge the (expected) *relative* performance of proposed RAS solutions against current human-centric / manned solutions.

3.1 Effectiveness

Speed. Rapid action and the element of surprise are integral to defy counter measures and supersede adversaries. Improvements to reaction times, speed in decision-making and rapid mobility and deployment across land, air and sea would invaluably enhance the strategic position of the RNLA.[11] With the help of artificial intelligence, which stimulates rapid decision-making and prioritization of threats, RAS are already capable of surpassing human reaction times and shortening the OODA (Observe, Orient, Decide, Act) loop.[12]

Speed is also an important metric of evaluation when considering 'Search and Support' roles of RAS, particularly maneuverability across conflict spaces. With logistical support from RAS, operatives can move to, from and around conflict spaces at greater ease, with less physical load and subsequently, at a faster pace.

Reliability. Delegating tasks to machines requires an immense degree of trust, especially considering the critical situations that the military is designed to thrive in. A key element of this trust is reliability, i.e. *has this function consistently worked effectively in previous circumstances?* When given critical situation, command would choose a human operative they could rely on over a machine that could do the job better - but only sometimes. At present, RAS are yet to prove adequate reliability across all military application areas. However, just as we trust a GPS over our own senses of navigation, so too will our confidence in other technological systems increase as they prove their reliability and effectiveness in executing specific tasks.

Accuracy. Accuracy is particularly contentious when it comes to strike capabilities as militaries seek to diminish the collateral damage of conflict and adhere to international standards on the protection of civilians. One of the main arguments in favor of unmanned aircraft strike (drone) missions, is that it is more precise. However, while AI systems have developed facial image recognition and sensory abilities past the level of human performance, the claim that unmanned systems are more precise than human operatives is widely disputed. A 2016 study disproved the claim that unmanned drones are more 'precise' and cause fewer civilian fatalities than airstrikes by manned aircrafts.[13] In fact, the research found that drone strikes are approximately thirty times more likely to result in a civilian fatality than an airstrike by a manned aircraft.[14]

[11] UK Ministry of Defence, "Joint Concept Note 1/18," 13.
[12] Reilly, "Beyond Video Games."
[13] Wolf and Zenko, "Drones Kill More Civilians Than Pilots Do."
[14] Wolf and Zenko.

Mass. Owing to increased range and endurance, RAS has the capability to enhance coverage of the battlespace and overwhelm adversaries. The best example of this potential is swarming, whereby a large quantity of physical, multi-robot systems use AI and advanced network communication to conduct highly-coordinated operations. With this coordination and smart mass, swarms are able to apply sustained pressure and use the frenzy of a simultaneous offensive to overwhelm adversaries. Additionally, whereas traditional mass involving concentrated force is problematic in terms of coordination and concealment, and dispersed systems are vulnerable to deficiencies in command and control, RAS have the potential to combine firepower, coordinated control and maneuverability.[15] Therefore, small systems but with superior AI will have the ability to defeat systems that use more traditional force. The risk here is that if communication between the multiple robotic systems is cut off (i.e. due to signal jamming), an affront ceases to be a concentrated and concerted effort, and is subsequently rendered futile. However, with continued advancements in military technology (including the resilience of device-to-device communication systems) it is plausible to see a shift from greater physical mass towards smart mass.

Reach. Similar to mass, extending reach is highly dependent on the ability of cells within the system to communicate and coordinate. When compared to human combatants, RAS greatly enhance the available points of presence for surveillance, intelligence, reconnaissance and weapons systems. This pertains not merely to the scope of the physical battlespace, but also to the use of RAS in cyber operations.[16] Furthermore, the use of RAS for Service and Support can extend the reach of human operatives on the ground by prolonging the moment whereby fatigue, diminishing supplies and transport maintenance restrict the length of a mission.

Robustness. The quality of being strong and/or unlikely to break or fail, especially during unexpected circumstances and against shocks, is particularly important given the hazardous nature of the operational environment. The development and implementation of quality assurance standards and certification processes will be critical in this regard. At least in the short term, RAS will be more vulnerable than humans to fail due to a small detail or an unanticipated change to the mission. This frailty extends beyond the physical domain towards the virtual domain, as losses in connection (for example through signal jamming), hacking and other interference can render a system incapable.

Safety. A distinct advantage of the integration of RAS into military functions is their ability to perform 'dull, dangerous and dirty' tasks. This leaves humans to focus on the more specialized tasks instead of those which are repetitious and messy, and most importantly, to be kept out of the line of fire. Although it is undeniable that remote controlled robots are saving the lives of soldiers, the strong emotional bond that humans form with their robotic team members can, in exceptional circumstances, have a paradoxical effect as soldiers have been known to risk their lives to save robots.[17] Aside from this, as advancements in robotic systems and human-machine teaming

[15] UK Ministry of Defence, "Joint Concept Note 1/18," 34.

[16] UK Ministry of Defence, vi.

[17] Hsu, "Real Soldiers Love Their Robot Brethren."

continue, and the technology gains trust through reliability, their use in more dangerous missions will intensify and we can subsequently expect a greater degree of safety for troops.

3.2 Efficiency

Cost. Efficient use of RAS has potential to substantially increase the cost effectiveness of defense processes.[18] Currently, the cost associated with pioneering and/or obtaining the latest RAS means that the development of this technology has been undertaken by a relatively small group of actors. However, as developers learn to adapt the technology in commercially available systems (such as smartphones), capabilities such as image recognition, navigation and remote operation, will become far less costly to acquire. Although exclusive access to the most cutting-edge technology will be reserved for the wealthiest players, the cost of systems that are now considered highly-advanced will fall throughout the next twenty years, thus becoming more widely attainable.[19] The degree to which RAS technologies are cost effective is also highly dependent on other evaluation metrics, such as whether the system is agile and applicable to multiple scenarios.

Maintenance. As with any technology, RAS require software and hardware upgrades in order to sustain accelerated capability development. Maintenance may also come in the form of fixing existing systems. While this metric is especially difficult to evaluate for RAS in general, it is nonetheless an important factor to consider when developing, purchasing or introducing RAS into a context.

Time efficiency. The performance of RAS in regard to time efficiency is one of the strongest arguments in support of its deeper and more widespread integration into militaries. RAS can perform dull and repetitive monitoring tasks at a high standard 24/7 without the need for rest, logistical planning can be solved efficiently, and the limits of human multitasking can be quickly surpassed.[20] This efficiency also allows for fast deployment and the reconfiguration of plans *en route*.[21]

3.3 Agility

Flexibility. Flexibility refers to the ability to change or be changed easily according to the situation. A flexible system can take on a variety of missions and/or perform these missions under a wide range of circumstances (e.g. climate, weather and terrain). Although RAS currently excel in executing specific tasks and humans will remain the most flexible for the foreseeable future. This dynamic is likely to change as developers continue to innovate current systems. Presently, RAS

[18] UK Ministry of Defence, “Joint Concept Note 1/18,” 7.
[19] UK Ministry of Defence, 5.
[20] UK Ministry of Defence, 44.
[21] U.S. Army, “The U.S. Army Robotic and Autonomous Systems Strategy,” 10.

extend the current flexibility of humans through human-machine teaming. An example of this is a service and support drone that can transcend the limits of a human team's ability to surveil in harsh environments, such as deserts. Thus, when the mission encounters (unexpected) challenges, RAS have the capability to make the team more flexible.

Adaptiveness. By contrast, adaptiveness indicates how a system may be changed over time or according to new circumstances. Where flexibility pertains to the versatility of the system as-is (i.e. during a mission), adaptiveness considers the potential of the system to be easily reconfigured (scaled, extended, upgraded etc.) over time so to keep up with new requirements emerging in a dynamic environment (i.e. during the system's life cycle).

3.4 Legitimacy

External. Legitimate use of RAS encapsulates compliance with the Dutch Constitution, international law (including Laws of War) and national legislation. Acting in accordance with the law becomes more contentious with higher degrees of autonomy and lethality of RAS, as discussed earlier. The establishment of certification regimes and clarification on precisely how (international) law applies to the development of RAS is instrumental in evaluating legitimacy in this regard. Additionally, as the army seeks to be engaged in society and inherently reflects the values of the society it serves, it must act within the parameters of societal acceptability. While these parameters are fluid and illusive in a continuously evolving society, positive public opinion (or at least passive acceptance) is of great importance to the army. As a socially responsible organization, RNLA's engagement with RAS must thus strike a balance between the advanced capabilities they (potentially) provide and the values and norms of the society it serves.

Internal. As operators of RAS, the RNLA must also be willing to implement RAS into their operations. This willingness is not only dependent on the external legitimacy of the system (legality, certification and ethics), but also on the degree to which the system is trusted to execute a task.[22] Trust and organizational normalization of RAS will be strengthened over time as understanding of the systems, their predictability and their familiarity are enhanced.[23]

[22] UK Ministry of Defence, "Joint Concept Note 1/18," 48.
[23] UK Ministry of Defence, 48.

4. Current RAS applications in the land domain

4.1 Overview of current systems

The dataset of RAS used by HCSS largely builds upon a SIPRI dataset which encompasses over 380 RAS classified into a number of general categories.[24] Our overview currently comprises 299 distinct RAS solutions. The majority of RAS are categorized under **Information and Intelligence**, while **Use of Force** has the least RAS solutions. It might be that the actual amount of RAS for **Use of Force** is broader than what can be asserted, precisely because of limitations due to the classification of matters concerning national security. Furthermore, collecting data on RAS in countries such as China and Russia is restricted by their known secrecy as well as language barriers.

In this part, a factual overview of current RAS is depicted, using the HCSS taxonomy of military functions (see *Figure 1*). The section will proceed by firstly demonstrating the first tier of this taxonomy; and then by the second tier, which offers a more detailed account of potential military applications of RAS. In furtherance of providing a clear broad view of RAS production and use, visualizations will display the approximate number of projects produced/ employed per country.

We categorized 299 RAS on the basis of their military functions, namely in the domains of **Service & Support**, **Information & Intelligence**, **Defensive Use of Force** and **Offensive Use of Force**, forming the first level of categorization. *Figure 3,* portrays the second tier of RAS, offering a more comprehensive view on HCSS' taxonomy of the systems.

[24] Despite the comprehensiveness of the SIPRI list, it contains several limitations, in particular with regards to its generic classification of RAS based on their purpose, i.e. their function. The SIPRI dataset ranges from systems that are operational, under development and cancelled/retired. For our purposes, the systems which are either retired or cancelled were excluded, along with the systems employed in the maritime domain. Our overview also made use of systems identified in the British Army Innovation Technology Book (BAITB), as well as studies conducted by the U.S. Congressional Research Service, Zhifeng Lim, and Boulanin & Verbruggen. The resulting dataset used by HCSS is for some 90% based on the SIPRI dataset, for 4-5% on the British Army Technology book, and for 5-6% on the additional studies. The data presented in this paper is accurate as of time of writing: March 2019. "Army Warfighting Experiment 2018: Autonomous Warrior"; Feickert et al., *U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI)*; Lim, "The Rise of Robots and the Implications for Military Organizations"; Boulanin and Verbruggem, "Mapping the Development of Autonomy in Weapon Systems."

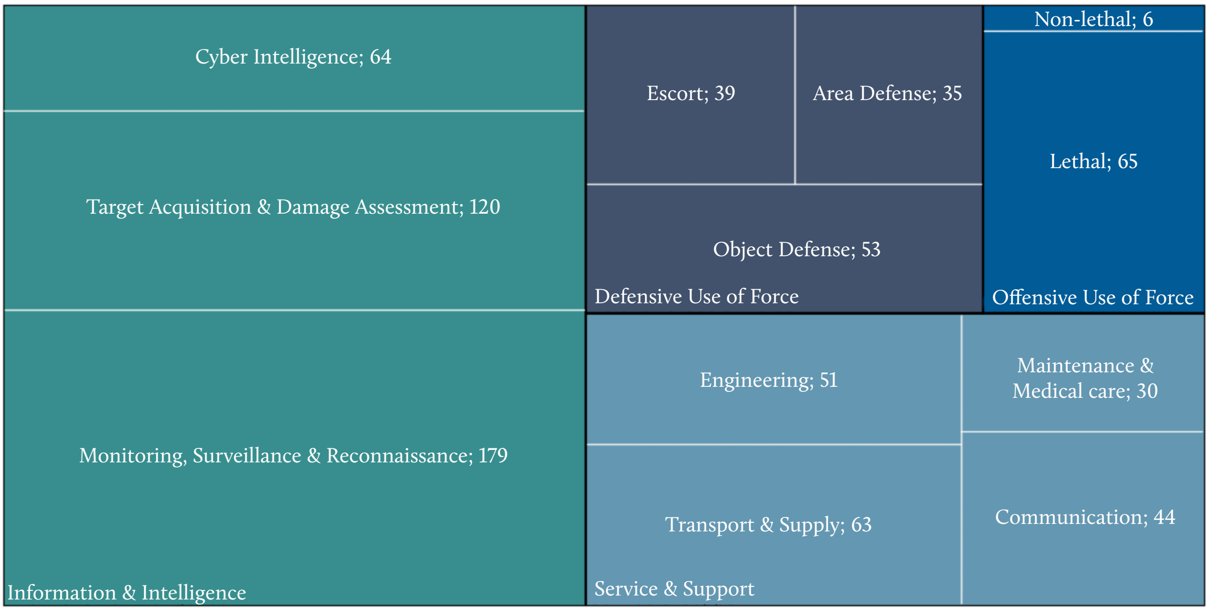


Figure 3. Represents Tier 2 of The HCSS RAS Taxonomy

As exhibited below in *Figure 4,* the majority of RAS are used for **Information and Intelligence** gathering purposes, with a total of 224 Systems, including *Hermes 900*[25] and *Nerva.*[26] The second most prevalent domain is **Service and Support**, encompassing 126 RAS, such as the *Amulet UAS*[27] and *Guardium-LS.*[28] In regards to use of force, 85 RAS were labeled under **Defensive**, with systems alike *Otomatic*[29] and the *Norwegian Advanced Surface-to-Air Missile System (NASAMS).*[30] Lastly, 69 systems were recognized as **Offensive**, and those include *Skystriker.*[31]

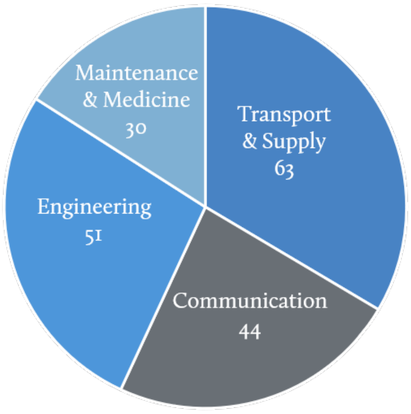


Figure 4. Categorization of Tier 1

[25] *Hermes 900* is an unmanned air vehicle (UAV) system that is used by the Israeli Defense Force for operations requiring intelligence, surveillance, target acquisition and reconnaissance (ISTAR).

[26] *Nerva* is a 2-wheel compact robotic platform, equipped with high-definition and thermal camera to serve its reconnaissance purpose.

[27] The *Amulet* is an unmanned air system (UAS) that is able to detect buried landlines, improvized explosive devices (IEDs) and emplaced explosive ordnance from a standoff distance.

[28] Guardium-LS is a multi-purpose autonomous unmanned ground vehicle that is able to provide troops with 1.2 tons worth of ammunition and supplies without endangering manned vehicles over routes stricken with IEDs.

[29] *Otomatic* is an armored anti-aircraft vehicle with the ability to detect enemy stealth aircraft.

[30] *NASAMS* s a medium to long range air defense missile system. It can recognize, engage and destroy helicopters, aircraft, cruise missiles and UAVs, and protects against air-to-surface threats.

[31] *Skystriker* is a Loitering Munition (LM) designed for use by the tactical level corps. The LM is able to seek, target and engage various targets.

Information & Intelligence. This category is branched into **Monitoring, Surveillance and Reconnaissance (MS&R)**, **Target Acquisition and Battle Damage Assessment (TA&BDA)** and **Cyber Intelligence**. *Figure 5* pinpoints MS&R as the largest sector with 179 RAS, TA&BDA is consequent at 120 systems, and finally, 64 under Cyber Intelligence.

Service and Support. *Figure 6* exhibits the **Service and Support** category with its four sectors, namely **Transport and Supply**; **Engineering**; **Communication** and lastly, **Maintenance and Medicine.** As the pie chart shows, the Transport and Supply sector is leading with 63 systems, followed by Communication at 44, Engineering at 51 and lastly Maintenance and Medicine covering a mere 30**.**

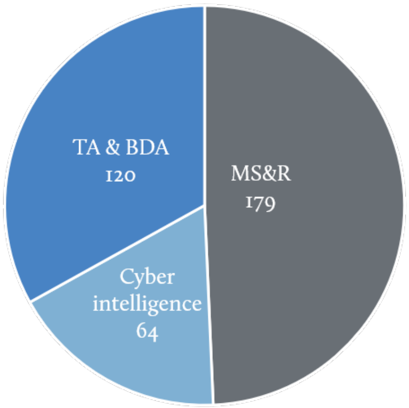


Figure 5. Information and Intelligence Systems

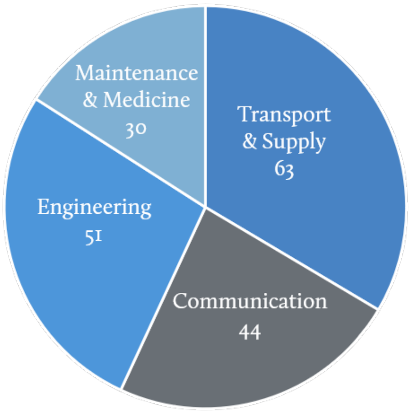


Figure 6. Service and Support Categorization

Defensive/Offensive Use of Force. In regards to **Use of Force**, *Figure 7* displays the **Defensive** RAS, which are divided into **Object**, 55 systems, **Escort** 39 systems, and **Area** 35 systems. Defensive RAS are notably more used for the purpose of Object protection, rather than Area or Escort protection.

On the other hand, *Figure 8* mirrors the **Offensive** RAS, grouped into 65 systems identified as **Lethal** and 6 systems as **Non-Lethal**. A significant discrepancy can be observed between the small amount of non-lethal RAS, in comparison with the amount of RAS that are used for lethal purposes.

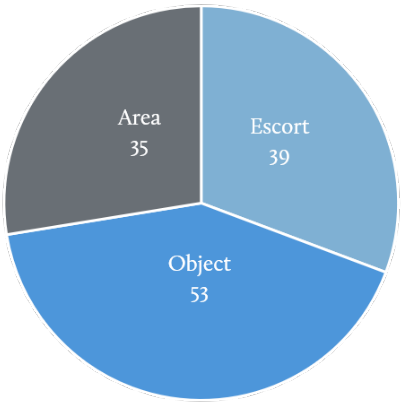


Figure 7. Defensive Use of Force Categorization

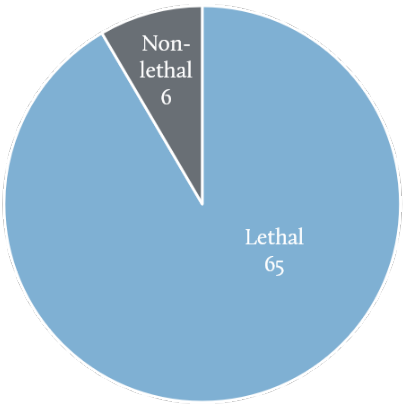


Figure 8. Offensive Use of Force Categorization

4.2 RAS per Country of Origin

Figure 9 visualizes the amount of RAS produced on a global scale. The chart outlines the number of RAS originating from various countries. These include, in a descending order, USA (81), Israel (48), Russia (33), Italy (25), France (23), India (21), China (19), UK (15), Germany (13), South Korea (11) and a cluster of countries in which 10 or less RAS are developed.

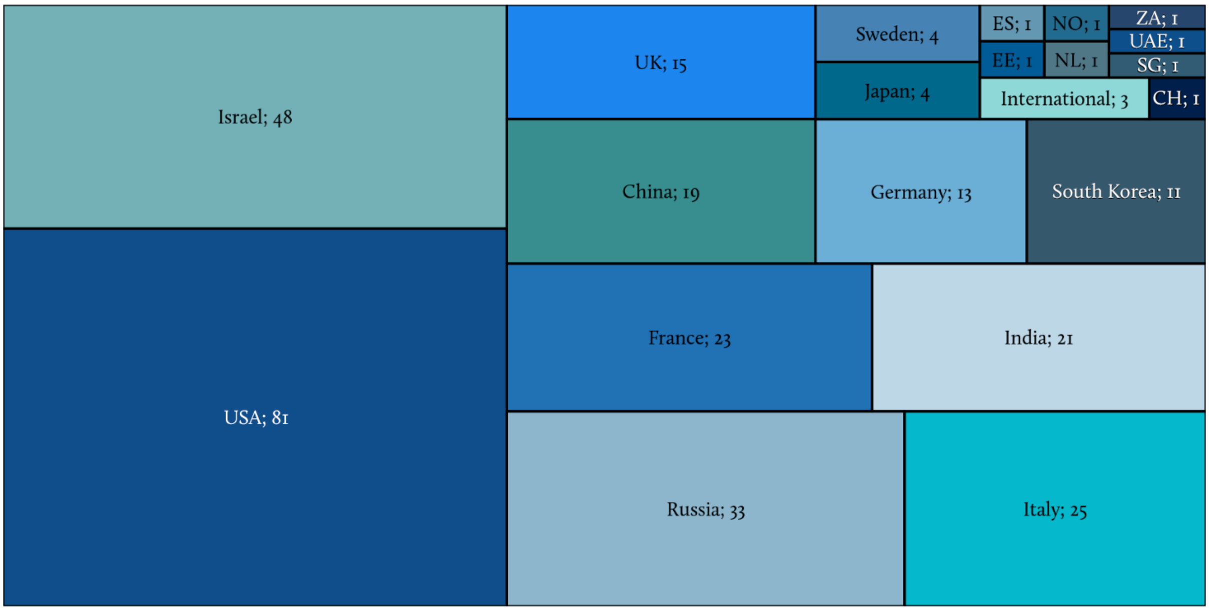


Figure 9. Number of RAS produced per Country of Origin

4.3 RAS per Country of Use

In an attempt to offer a coherent overall view of the global use of RAS, *Figure 10* provides a map presenting the countries that make use of RAS. Additionally, countries were color characterized on the basis of volume: 51-100 projects are marked in red, 21-50 in orange, 11-20 in yellow and 1-10 in purple.

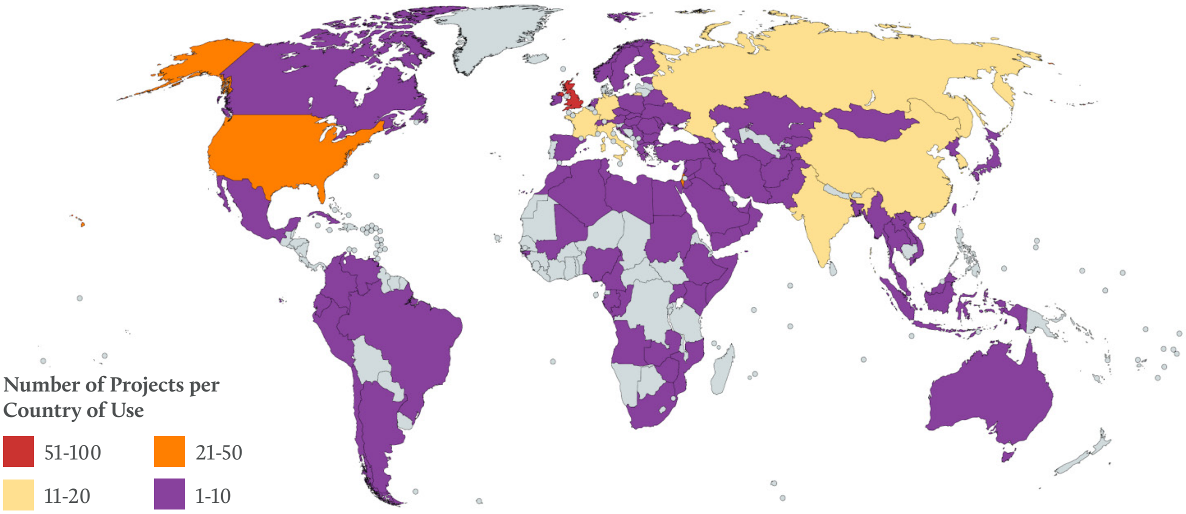


Figure 10. Number of Projects per Country of Use

4.4 Future Applications (Military applicability of RAS for the next 5-10 years)

This section provides an overview of the application areas where RAS could be potentially introduced into the armed forces, with a particular emphasis on land forces in the next five to ten years. Categorized according to tier one of the taxonomy, the findings provided below are based upon the results of an expert workshop conducted by HCSS and the RNLA, combined with an open-source research.

Service and Support

With regards to the service and support application area, workshop participants noted that RAS could assist or augment a number of human centric solutions, especially in the domains of level 2 of the taxonomy (i.e. transport and supply). For instance, a system used for supplying military equipment, as well as medical evacuation and convoy protection, could provide solutions to tasks that are often considered too dangerous for a human to perform. An example of such a system is the U.S. Crusher UGV. Capable of carrying over 8000 pounds of payload at high-off road speeds and across extreme terrains, Crusher can provide increased mobility, reliability, and logistical support for the army personnel on the ground.[32]

RAS could also be expected to perform various engineering functions, such as C-IED, laying and building bridges, and repairing and maintaining military equipment. When it comes to tasks that have no human-centric alternative, participants pointed out that RAS could be used as mobile communications relay stations and as autonomous recovery systems for RAS. Instead of dispatch manned personnel, such systems could recover damaged equipment, vehicles, and so forth in terrain that is considered dangerous. In this regard, the utility of RAS that are capable of self-healing, self-assembling, and self-repairing is expected to be extremely high in the future to come.[33]

Additionally, RAS can offer effective C3, thus, allowing for the further centralization and increased effectiveness of military command and communication. An example of a system that is capable of performing such functions is the Ground Control System (GCS), identified in the British Army Technology Book. GCS is a platform that enables optimal flexibility between airborne and ground systems, allowing for a more effective command and control during ISR operations.[34]

Information and Intelligence

In the Information and Intelligence domain, workshop participants indicated that RAS could provide the most military utility in functions such as perimeter monitoring and control—for both in missions abroad and as base protection at home—TA&BDA, and enemy deception. In regards to the latter, a particular example was identified: a system that could deceive the enemy by creating artificial objects used for visual deception. A system of such type is observable in Israel's *Project Hyen*a, which infuses sounds and signatures of real tanks and other vehicles with foldable, semi-autonomous lightweight platforms in order to deceive and provoke enemy objects.[35]

It was also highlighted that RAS could be expected to augment human centric solutions in providing advice to the army by observing, recognizing, and analyzing information autonomously. For instance, *Mantis* UAV allows the UK Defense Forces to perform ISTAR operations, while offering close-air support for ground missions as well as capturing and transmitting , via satellite link, real-time data to the ground control station.[36] Furthermore, in terms of tasks with no human-centric alternative, participants underlined the importance of RAS in providing tactical ISR in urban terrain, i.e. inside of buildings and other urban premises, underground areas, and so forth. The Casper 250 backpack mini UAV is one of the systems that can be operated for this purpose.[37]

[32] National Robotics Engineering Center, "Autonomous Platform Demonstrator."
[33] Nathan Fisher and Gary Gilbert, "Medical Robotic and Autonomous System Technology Enablers for the Multi-Domain Battle 2030-2050."
[34] "Army Warfighting Experiment 2018: Autonomous Warrior," 10.
[35] Defence Industries, "Https."
[36] Airforce Technology, "Mantis MALE Unmanned Aerial Vehicle (UAV)."
[37] "Casper 250."

Defensive Use of Force

In regard to defensive use of force, workshop participants underscored the importance of RAS in providing electronic countermeasures as part of area/perimeter protection. The Israeli UGV, Loyal Partner, constitutes such an example which can respond to suspicious attacks and eliminate various threats by using forceful methods, thus protecting manned troops on the ground.[38] RAS could also be defensively deployed in cases of terrorist and/or militia attacks. Participants pointed out to a particular example: a hybrid system—operating in both land and air domains—that would have the ability to attract enemy fire, thus protecting the lives of military personnel on the ground. The HIPPO All Terrain Support Vehicle (ATSV) is a relevant example of a system that can contribute in performing such tasks. With the ability to perform surveillance functions and carry weapons, HIPPO ATSV can provide immense strike effects whilst minimizing the exposure of manned personnel to enemy fire.[39]

Offensive Use of Force

In the offensive use of force application area, workshop participants recognized the ability of RAS to replace traditional jet fighters, such as F-16 and F-35, with sophisticated lethal drones that in the future could fly in swarms. Such a system that is currently under development is the U.S. X-47B Unmanned Combat Air System (UCAS). About the size of an F-16, yet faster and lighter, this UCAS is capable of performing full-fledged jet fighter operations.[40] Additional RAS identified in this application area constitute loitering munitions, and an anti-tank vehicle. In regards to the former, a system that possesses the potential to be applied in the future is the British HERO 30 Tactical Precision Missile System. A lightweight lethal missile system whose preparation and silent launch take less than two minutes, HERO30 is capable of destroying targets at up to forty kilometers in range.[41] In relation to an anti-tank system, the Russian Platform-M UGV constitutes a relevant example. Equipped with anti-tank rockets, Platform-M UGV is capable of performing a wide-range of combat tasks during both day and night with no need to unmask itself.[42] Finally, participants recognized that in terms of tasks with no human-centric alternative, in the future RAS could provide counter-A2/AD measures primarily through saturation tactics.

[38] Army Recognition Group, “Ground and Aerial Unmanned Vehicles UGV UAV in the Israeli Army Defence Forces.”

[39] “Army Warfighting Experiment 2018: Autonomous Warrior.”

[40] Smith, “The U.S. Navy Spent $744 Million to Build a Robotic Fighter Jet -- and Now Wants to Throw It Away -.”

[41] “Army Warfighting Experiment 2018: Autonomous Warrior,” 92.

[42] Army Recognition Group, “Russian Special Forces Have Received Platform-M UGV Unmanned Ground Vehicles.”

5. Opportunities and Risks

RAS presents numerous, significant and far-reaching opportunities for the RNLA. To summarize, this includes creating better and faster situational awareness and understanding, reducing the physical and cognitive loads of soldiers, sustaining and protecting the force, extending the reach and persistence of operations, increasing the pace of the OODA loop, and allowing the simultaneous execution of tasks for efficient action. Across all domains, the current human-centric boundaries to speed, knowledge, endurance, scale, accuracy and flexibility will be pushed forward to new, ever-expanding limits. However, with these opportunities come significant challenges, both in terms of practical issues within the military and also in terms of conditional, external issues.

5.1 Practical (Internal) Challenges

Technical

- The signal communications used by RAS are vulnerable to cyber attacks, including hacking, jamming, 'spoofing', or otherwise impeding the performance of the system. Commercially available software and hardware is capable of achieving these effects.
- In order to facilitate trust in decision making, operators must understand how the system interprets data and delivers actionable information, which can be extremely difficult for highly complex systems.
- Currently, there is a lack of understanding within defense communities of how to operate RAS and how to fix (minor) problems that arise.

Personnel

- There is currently not enough trust in RAS, especially in high-stake situations, thereby impeding the advancement of man-machine teaming to its full potential.
- Operators may develop overconfidence and/or overdependence on RAS.
- Operators are susceptible to becoming mere acceptors of the 'decisions' made by RAS, without oversight of the algorithmic processes preceding the outcomes. The degree to which humans are meaningfully present in the OODA loop is therefore questionable.
- As new technical experts and data scientists are required and recruited, internal tensions between new technical personnel and traditional soldiers may arise.
- Organizational Culture.

- RAS will certainly lead to changes in terms of training requirements, education, careers and the type of work soldiers engage in. Leadership positions will also change in character.
- Experimentation and rapid innovation do not align with a culture of meticulous planning and linear requirements assessment, development and acquisition processes.
- Inefficient procurement processes can lead to difficulty in keeping up with the speed of technical advancements, in particular for technology developments that are driven by commercial markets.

Doctrinal

- The integration of RAS and the possibility that machines will replace humans or units in some capacity, will have implications for the doctrine of the Dutch armed forces, but also to the doctrines of allies, such as NATO.

5.2 Conditional (External) Challenges

Perceptions

- Due to the perpetuation of the "killer robot" image, public perception appears generally negative, despite the nuanced nature of the situation. This dystopian imagery may lead to the requirement of human control across all application areas, regardless of the benefits of automation in most cases.
- It is possible that the use of LAWS will create anti-Western sentiment (and even radicalization) in the areas affected by strikes, due to the perceived indignity and unfairness of being injured or killed by an unmanned system.

Ethical

- Ethical discourse on RAS focuses on the interplay between offensive and defensive use of force, the necessity of 'meaningful human control' and the question of human dignity.
- There is also an insinuation that RAS may lower the threshold for escalation of conflict due to the dehumanization of the use of force.
- Despots and rogue states, who are less concerned about ethical considerations, may proliferate RAS and more ethical states could subsequently 'fall behind'.
- The acceptance of adverse effects is in proportion with strategic interest. The benefits of broad implementation of RAS may be a higher priority than ethical consideration, and political discourse can reflect this message.

Legal

- In terms of legality, it is not yet clear exactly how international and national law will adapt to - let alone anticipate - this rapid, and exponential technological change.
- Attribution will become a growing challenge as actors who use RAS will be increasingly able to deflect or avoid responsibility for attacks.[43]

Military

- Adversaries may face few ethical and legal limitations to the proliferation and use of RAS across all domains.
- The cost of systems that are now considered to be elite will fall. These systems may subsequently be acquired by smaller actors, including non-state actors.

Political

- RAS represents the next revolution in warfare and powerful states are racing to harness the potential. This will likely lead to an arms race.
- While the state retains a monopoly over the legitimate use of force, other actors such as private military companies, paramilitaries and non-state actors (each with their own agendas) will be involved in the procurement of RAS.

6. Next Steps

These opportunities and challenges imply prerequisite measures for the successful implementation of RAS, as well as ways to mitigate (potential) challenges. Based on knowledge of the current RAS applications in the land domain, the assessment of the military value of RAS, and the identified opportunities and challenges that were identified in the workshop, the following measures were observed by the practitioners, researchers, legal specialists, ethicists, members of the defense community, industry professionals, academics and researchers present during the Expert workshop.

These measures represent the insights of this community, derived from a dynamic and inspiring group process during the workshop, on how to proceed with the development of RAS in the land domain. These insights should be used as inspiration for developing strategies and policies, and as basis of knowledge for the next essential steps to be made in the public and political debate on RAS. Legal, ethical and other debates with regard to RAS can only be done sharply if there is a common and good understanding of what the military application of RAS in the land domain means and what is necessary for the inevitable and essential development of this military technology.

The measures are divided into two categories: what should shape the strategies and policies within the army and how to align with other communities. The measures are not prioritized and are all essential in the development of RAS in the military context.

6.1 Within the RNLA

- In order to reap the full potential added value of RAS, concept development & experimentation (CD&E) within the armed forces is urgently needed. This should be done in 'open' configurations, because most of the relevant technology and technology development should be crossed-over from non-military sectors and application areas.
- As the unrestricted use of the cyber and electromagnetic domain is a key requirement for the use of all new technologies in the information age, cyber security and electromagnetic spectrum security, as well as control over the electromagnetic spectrum (in a 'battle over bandwidth' with adversaries), are top priorities to ensure the integrity of RAS. In addition to monetary investment, this requires mandating developers and producers of RAS products to prioritize security and connectivity, actively ensure that their products are free from vulnerabilities, and take timely action to mitigate vulnerabilities that are later discovered.

- Financial investment into AI and systems engineering is vital for the advancement of capabilities such as speed, accuracy, reliability and flexibility. Throughout the development of RAS, there must be a focus on user friendly interfaces with adequate, human-centric checks and balances. During this interim period, it will be necessary to have an online, real-time operator helpdesk to aid operators if needed.
- Education and training programs must address the development of required skills within the organization. The high quality staff needed to develop and maintain RAS (e.g. IT specialists, software engineers), are in short supply and are often lured by attractive salaries in the private sector. Therefore, the RNLA needs to take steps to develop an organizational culture and individual mindset of continuous learning and improving. For the organization, this not only requires education and training of its current employees, but also continuously searching, identifying, contracting and training new people.
- The training, improvement, and effective maximization of AI requires access to large, high-quality datasets. This requires not only a new kind of knowledge and the implementation of new technologies, but especially highly skilled specialists. Skilled people are the most scarce capability and to attract and keep them on board, a partnership approach is needed.
- A partnership approach, instead of competition, needs to be developed between the private sector and the land forces to facilitate decentralized innovation and CD&E. The RNLA and the Ministry of Defence (MoD) need to adapt in order to make the new Army possible. Organizational culture needs to orient toward innovation and transformation by developing entrepreneurial spirit and facilitating more civil-military collaboration, multi-disciplinary interaction and exchanges of people and knowledge.
- The procurement processes of the MoD must also be adapted toward buying smart systems . This involves a more generous interpretation of European procurement rules and the acquisition of commercial 'off-the-shelf' RAS-capabilities that can be tailored (in close cooperation with industry and knowledge institutes) to military applications. The speed of technological developments in this field requires continuous modifications of systems and implementation of new technologies. Procurement processes need to facilitate this new, permanent Beta approach.
- In order to develop the trust of operators, the system must be predictable, familiar, understandable and appropriate for the context in which it is operating. On a higher level, certification regimes must be established, promoted and implemented. The RNLA has agency in shaping responsible norms around verification of producers and quality assurance of RAS products.

6.2 External to the RNLA

- More understanding and information needs to be disseminated to the public and policy makers regarding the development and use of RAS across all application areas. Ethical standards should be derived from a rich and well-informed debate. The MoD must be active and visible within this national conversation, and transparent in their actions and intentions. Positive aspects of

RAS should be emphasized alongside the challenges. The dystopian image of RAS will otherwise lead to the necessity of human control across all domains.

- The MoD must observe and strategically account for the growing capabilities of RAS systems in foreign states. This insight must be debated, politically and publically, and periodically translated into the military strategy for the development of the army. Furthermore, these insights need to be shared with partnering countries and synchronized with the military strategies of these partners. In our current, developing, globalized and multipolar political system this becomes ever more relevant.
- International legal regimes must work to develop explicit and simple guidelines on autonomy across all application areas, particularly in regard to LAWS. This requires a sharp insight into the targeting cycle of LAWS and the different steps within this process. Added knowledge in this area will bring the focus of the debate to those steps of the targeting cycle that are most contentious. The state of technology and the experiences humans have with these technologies determine the trust in these new technologies and the level of automation or human control. Therefore, the debate on autonomy in military applications will continuously develop.

Factsheet

The Ethics of RAS in a Military Context

The increasing importance of RAS

The accelerating arms race in Artificial Intelligence (AI) and the diffusion of cheap but technologically advanced military systems among state and non-state actors compels militaries to adopt robotic and autonomous systems (RAS). How can countries stay competitive, without violating core national and international ethical principles?

Key Ethical Challenges of RAS

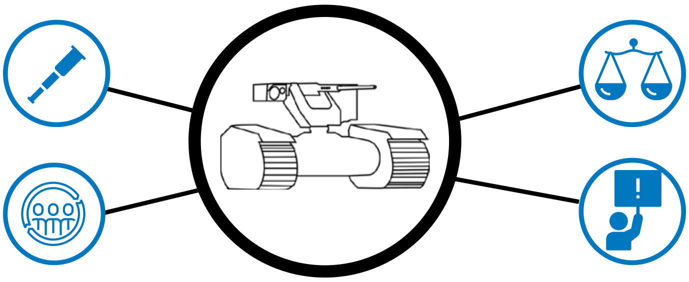
- The ability of humans to retain control over systems
- Respect for human dignity
- The shortcomings of present responsibility structures

Human Agency

The establishment of meaningful human control over RAS should be preceded by the identification of the type of autonomy in the system and the function it serves.

Human Dignity

Crucial to upholding human dignity in conflict is adherence to International Humanitarian Law: RAS should be able to respect proportionality, military necessity, distinction and humanity.



Robotic and Automous Systems (RAS)

RAS is an all-encompassing term that refers to systems with any degree of autonomy and any military designation, whether for communication, logistics, reconnaissance or weapons delivery.

Responsibility & Accountability

Accountability should be addressed at an institutional level all throughout the life cycle of a system. Responsibility is fragmented across many actors - therefore it's crucial to ensure behaviour can be accounted for.

Recommendations for the Royal Netherlands Army

- Adapt internal processes to address technological developments
- Develop guidelines for identifying use-cases and developing ethical RAS
- Consider studying RAS-testing approaches of other states
- Identify best practices for military-private sector cooperation
- Communicate research into and use of RAS to the public
- Ingrain an institutional approach to the responsibilities of all actors involved
- Continue to research the role of RAS in the military context

The Hague Centre for Strategic Studies

The Hague Centre
for Strategic Studies
Lange Voorhout 1
2514 EA The Hague
The Netherlands
hcss.nl

Chapter 2

The Ethics of Robotic and Autonomous Systems in a Military Context

Esther Chavannes & Amit Arkhipov-Goyal

Executive Summary

The current ethical debate on robotic and autonomous systems (RAS) is often dominated by relatively extreme narratives akin to banning ‘killer robots’ (a euphemism for lethal autonomous weapon systems) entirely. While many ethical concerns are substantive, discussions on RAS have side-lined nuances that have critical implications on deciding how to introduce RAS in a military context. The brewing AI arms race and the diffusion of cheap, technologically advanced systems among state and non-state actors compels countries to adopt RAS, due to the prospect of lagging behind allies and adversaries are likely to work on RAS development. With the perspective that RAS will be further incorporated in the military context, this paper presents a balanced discussion on the key ethical challenges arising from the introduction of RAS; human agency, human dignity, and responsibility. In short, these topics of debate 1) concern the ability of humans to retain control over systems; 2) weigh the positive and negative ways in which RAS in a military context contribute to respect for human dignity; and 3) assess the shortcomings of present responsibility structures for deploying RAS.

(Semi-)autonomous systems have been in operation for over four decades and have previously received relatively little ethical consideration. As systems become increasingly independent with the capability to perform their own calculations rather than just being bound by a set of rules, the concern for rogue robots has arisen. In reality, however, the advent of systems matching human intelligence is unlikely to be achieved in the following decade. This redirects the focus to more functional challenges, such as the design of systems, decreasing understanding of algorithmic

calculations, and cognitive challenges arising from human–machine teaming. To address this, the paper presents a three-part framework through which to identify human control within a system: through the life cycle of RAS, through their sub-system functions, and through the observe-orient-decide-act (OODA) loop.

The paper discusses the impact of RAS on human dignity in the form of a debate between arguments that the use of RAS may aid or may undermine respect for human dignity. The ethical debate on human dignity is, for the purposes of clarity, prefaced by a detailed discussion of the key obligations for militaries in conflict, embodied by International Humanitarian Law. Hence, the collective decision to deploy RAS is that of the public, the government and the armed forces, but one that must be made on an informed basis.

The final part of the full ethical overview of RAS is the topic of responsibility. This is not only important for cases of active wrongdoing, but it is at least as crucial for identifying and addressing mistakes that may occur in the further integrating of RAS into the military. Aside from International Humanitarian Law, which can govern users', operators' and commanders' behavior with RAS in warfare, elements of both civil and criminal law may be relevant in addressing questions of responsibility and accountability for the actions of RAS.

As a result of the study, several recommendations are offered to the Netherlands Ministry of Defence and the Royal Netherlands Army (the complete list of recommendations is provided in Chapter 6 of the paper):

- The fundamental principle is to work with 'ethics by design', whereby ethical considerations are incorporated in the use-case identification, system design, validation, manufacturing, and testing processes, rather than solely in the operation stage of the system life cycle;
- Build understanding of the system performance and behavior through the involvement of end users in the design and testing stages, with the end goal for the operators, supervisors and commanders to be able to trace, understand and predict the system's decision-making process;
- Develop best practice guidelines both for the (1) outsourcing of the development process to external contractors and; (2) for interoperability frameworks with technologically advanced allied armed forces that co-deploy RAS;
- Identify within what sub-system functions of RAS increasing automation and autonomy will present benefits to the military without eliciting major ethical concerns, e.g., movement controls, sensory controls and computer vision;
- Program core rules of engagement (ROEs) with International Humanitarian Law principles embedded in system design, along with an open architecture to introduce mission-specific ROEs by mission command;
- Improve transparency on the use and contexts of use of RAS in the military domain with the general public.

While the recommendations presented above are not the sole solutions to existing ethical challenges, they do present pathways for the systemic incorporation of ethical principles in RAS within the Royal Netherlands Army. Rapid advances in computational power and data generation are paving the way for exponential growth in the sophistication of RAS, making this is a salient issue for both the Netherlands Ministry of Defence and the Royal Netherlands Army, as well as governments and military forces elsewhere. The above recommendations are therefore presented with a distinct sense of urgency.

1. Introduction

Throughout history, the invention of new military technologies has fundamentally changed how wars are fought.[1] The introduction of Robotic and Autonomous Systems (RAS) is no different and has led to renewed concerns over ethical issues associated with the use of new technologies in military forces. This is particularly salient in the context of autonomous weapon systems (AWS).[2] Fully autonomous and human-supervised autonomous systems have been in operation for several decades in over thirty countries. These have previously raised little ethical concern, even with their high degree of autonomy and often lethal designation, such as the Israeli Harpy and the US Tomahawk Anti-Ship Missile,[3] the latter of which was already withdrawn from service in the US Navy in the 1990s.[4] Their application is most frequent in cases where engagements supersede human decision-making and reaction times, such as with surface-to-air missile launchers and close-in weapon systems. The proliferation of (semi-)autonomous systems has been expanding exponentially, with at least 16 countries and several non-state actors, such as Hezbollah in Lebanon and Houthi rebels in Yemen, being in possession of armed unmanned aerial vehicles (UAV).[5]

1.1 Definitions

Throughout this paper, a differentiation is made between Robotic and Autonomous Systems (RAS) and autonomous weapons systems (AWS), according to the following basic definitions of RAS and AWS:

Robotic and Autonomous Systems (RAS)

RAS is an accepted term in academia and the science and technology community and highlights the physical (robotic) and cognitive (autonomous) aspects of these systems. For the purposes of this concept, 'RAS' is a framework to describe systems with a robotic element, an autonomous element, or more commonly, both.[6]

[1] Banta, "'The Sort of War They Deserve'?"

[2] Some scholars indicate that this balancing difficulty is less attributable to technology alone, but as much so to governments' choices, such as the development toward presuming some right to anticipatory self-defense, like in the US' drone campaigns. It could even be argued that this ethical discussion is not new per se, and mirrors the one concerning the development of air power in World War II, see Boyle, "The Legal and Ethical Implications of Drone Warfare."

[3] The Tomahawk Anti-Ship Missile (TASM) should not be confused with the Tomahawk Land Attack Missile (TLAM), which is still in service to this date and operates under a different set of parameters. Scharre, *Army of None: Autonomous Weapons and the Future of War*, 47–49.

[4] Scharre, *Army of None: Autonomous Weapons and the Future of War*, pages 47-49.

[5] Scharre, 102–3.

[6] The definition is borrowed in full from Feickert et al., "U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI): Considerations for Congress."

For the purpose of this paper, RAS is an all-encompassing term that refers to systems with any degree of autonomy and any military designation, whether for communication, logistics, reconnaissance, weapons delivery or otherwise, meaning the definition is inclusive of, but is not limited to AWS.

Autonomous Weapon Systems (AWS)

These are weapon systems that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that can select and engage targets without further human input after activation but are designed to allow human operators to override the operation of the weapon system.[7] Variation upon this US Department of Defense definition exists internationally, and characteristics such as 'intelligence', the possibility to learn or adapt, or a level of unpredictability are sometimes included.[8] Some, for example, consider the ability to search for targets by maneuvering intelligently through an environment to be a feature of an autonomous weapon system.[9]

1.2 Ethical Controversy in the Use of RAS

Arguments for the importance of RAS are numerous, and primarily encompass the ensuing technological arms race, diffusion of military power, and societal expectations of lower numbers of civilian casualties.[10] Concern for the adversarial development of RAS is one of the frequently cited reasons for a state's own development of such systems.[11] While normative actors such as the Netherlands do not seek to delegate absolute authority to machines,[12] adversarial state and non-state actors in possession of autonomous systems may gain a competitive advantage and, as a result, present a security risk to the Netherlands. Among state actors, this is developing into an 'AI arms race', where countries feel the need to develop AI-driven systems because other states are or may be doing so. The open-source nature of AI and robotic hardware makes such systems accessible to both individuals and groups, creating a whole set of security risks.[13] Beyond the technological and strategic competition, Western societies have also come to expect fewer civilian and soldier casualties in warfare due to technological advancement of the arsenals of nation-states. While individual civilian casualties resulting from drone strikes are questioned today, just 75 years ago during World War II, nation-states were carpet bombing cities, with hundreds or thousands of deaths resulting from individual air campaigns.[14] As the societal tolerance for civilian casualties decreases, the need for advanced systems of defense and precision-guided weapons becomes more apparent.

[7] The definition is borrowed in full from Feickert et al.

[8] Kania, "China's Strategic Ambiguity and Shifting Approach to Lethal Autonomous Weapons Systems," April 17, 2018; Ekelhof, "Autonome Wapensystemen: Wat We Moeten Weten over de Toepassing van Het Humanitair Oorlogsrecht En de Menselijke Rol in Militaire Besluitvorming," 194; Chairperson of the Informal Meeting of Experts, "Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)."

[9] Scharre, *Army of None: Autonomous Weapons and the Future of War*, 123.

[10] Scharre, 95, 117, 134.

[11] Jones et al., "Managing the New Threat Landscape: Adapting the Tools of International Peace and Security," 18.

[12] Netherlands Advisory Council on International Affairs, "Autonomous Weapon Systems: The Need for Meaningful Human Control."

[13] Jones et al., "Managing the New Threat Landscape: Adapting the Tools of International Peace and Security."

[14] Scharre, *Army of None: Autonomous Weapons and the Future of War*, 282.

The International Committee of the Red Cross (ICRC) has concluded from various opinion polls in 50+ countries that, when it comes to autonomous weapon systems, the most prominent ethical concerns are those regarding "loss of human agency in decisions to use force—decisions to kill, injure and destroy—, loss of human dignity in the process of using force, and erosion of moral responsibility for these decisions."[15] When put into a broader perspective—to also include non-weapon systems—the same ethical concerns remain relevant. These concerns touch upon questions of human agency, human dignity, and responsibility. Although some overlap can be found between these three overarching ethical issues, together they cover the most relevant concerns associated with the adoption of RAS by armed forces.

Considering the recent developments in RAS proliferation, the systems' utility for militaries worldwide means that RAS are already widely used, and capabilities will continue to be developed further. **Therefore, the objective of this paper is to present existing ethical challenges for the use of RAS, highlight the issues that have previously received little attention, and discuss pathways for the ethical integration of RAS in the Netherlands Armed Forces (RNLA).** This paper will not only consider lawfulness and lawful use of RAS under International Humanitarian Law but will also delve into the moral discussion on the use of RAS and human dignity, as well as questions of understanding, bias, accountability and responsibility. Within this wider discussion on the ethical, legal and social ramifications of these technological developments,[16] the structure of this paper is based on the following questions:

- How and to what extent can and should human agency, and human control in particular, be retained in the operation of RAS?
- What does it take to maintain a human dignity in the operation of RAS?
- Who is, or should be, responsible for the actions and outcomes of the use of RAS?

These guiding questions and the three aforementioned key ethical issues form the structure of this paper. Following Chapter 2 on the background and recent developments relevant to discussions on RAS, Chapter 3 covers the topic of human agency, and human control in the use of RAS in particular. It discusses a framework for how to assess the different aspects of what constitutes 'meaningful human control', as well as challenges such as explainability, self-learning abilities, and the complexities of human–machine teaming. Chapter 4 on human dignity positions the ethical debate on RAS, and on AWS in particular, in the context of existing frameworks of International Humanitarian Law. It also discusses to what extent AWS can undermine or enhance human dignity in a military context. Finally, Chapter 5 on accountability and responsibility presents the complexity of holding states and/or individuals responsible in the search for legal accountability,

[15] While the ICRC is just one of the stakeholders in the discussion on the ethics of RAS, and the results of opinion surveys have limitations, the ICRC has guided the humanitarian perspective of warfare throughout modern history. Following the Geneva Conventions, the ICRC remains the primary normative actor focused on maintaining humanity in warfare and, as a result, presents the fundamental humanitarian concerns arising from the use of RAS in a military context; see Davison, "Autonomous Weapon Systems"; International Committee of the Red Cross ICRC, "Ethics and Autonomous Weapon Systems," 21.

[16] Floridi, "What the Near Future of Artificial Intelligence Could Be."

and it addresses the phenomenon of an 'accountability gap' formed by the outpacing of laws and social norms by technological progress.

1.3 Research Process

The debates have yet to result in concrete guidance for tackling the ethical challenges of RAS integration that the armed forces are faced with. This paper sought to address this in order to guide further discussions on the introduction of these systems within the Netherlands Armed Forces. This paper draws its conclusions from an extensive literature review and an expert session hosted by *The Hague* Centre for Strategic Studies (HCSS) on June 13th, 2019. The expert session was based on four fictional scenarios (attached in Appendix A) that were created to test the cognitive boundaries on the ethical challenges posed by the use of RAS in particular contexts. The session involved military, legal, technical, and ethics experts from the Netherlands Ministry of Defence, Royal Netherlands Army, Netherlands Organisation for Applied Scientific Research (TNO), academic institutions, non-governmental organizations and the private sector. The participants were assigned into four groups, with each group completing all four scenarios. This way, four sets of perspectives were recorded for each scenario and were subsequently noted in the session summaries provided in Appendix B.

2. Background and Recent Developments

It is critical to note that most applications of RAS in the military setting are not for lethal designations, and span a variety of roles including surveillance, logistics, medical support, maintenance, communication and engineering.[17] It is estimated that out the known 500 RAS in operation today worldwide, 30% are designated for the use of force, within which 55% are used for defensive and 45% are used for offensive purposes. This means that 14% of all systems currently employed have a lethal offensive component, and are thus the primary focus of this paper.[18] Meanwhile, non-lethal systems and applications continue to demonstrate landmark achievements, such as the US Navy's Autonomous Aerial Cargo/Utility System, which autonomously determined an improvised landing zone and carried out an autonomous landing in 2014.[19] A year later, the US X-47B UAV conducted the first fully autonomous air-to-air refueling.[20] In the summer of 2019, the Dutch Army 13th Brigade trained with two THeMIS combat support unmanned ground vehicles (UGV) in Scotland, introduced as logistic support for deployed troops (see Figure 1).[21] It is evident that most (non-lethal) systems are introduced to support the armed forces without ethically motivated pushback. Thus, this paper extensively focuses on the ethical challenges posed by AWS and is hence angled towards addressing shortcomings. This is not to suggest that non-lethal systems have no ethical challenges, rather that the paper focuses on the issues that have been most contentious and have even resulted in a widely reported pressure group "Campaign to Stop Killer Robots".[22]

Similarly to RAS, Artificial Intelligence (AI) algorithms that power autonomous systems have been in development since the 1950s, and their increasing abilities are driven by a newfound capacity to collect, store and process mass amounts of various types of data.[23] It is therefore critical to stress that neither RAS, nor the AI powering it, are new. Rather, due to the advancements of computational power and the increased amounts of data ('big data') generated in the last decade, capabilities are now expanding far beyond initial abilities, and in turn beyond the levels of human cognition. Breakthroughs have been made in the use of various deep learning (DL) algorithms such as deep

[17] Torossian et al., "Paper on the Military Applicability of Robotic and Autonomous Systems," 15.
[18] Percentage calculations are based on: Torossian et al., 15–16. While some systems have distinct applications, others have multiple purposes, such as unmanned aerial vehicles capable of acting both as surveillance and weapons delivery systems, meaning this can affect the abovementioned figures.
[19] Scharre, *Army of None: Autonomous Weapons and the Future of War*, 17.
[20] JASON, The MITRE Corporation, "Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to DoD," 4.
[21] "Milrem Robotics Delivered Two THeMIS UGVs to the Dutch Army."
[22] "Killer Robots - Learn."
[23] Scott et al., "Modeling Artificial Intelligence and Exploring Its Impact"; Spiegeleire, Maas, and Sweijs, *Artificial Intelligence and the Future of Defense*, 31–39.

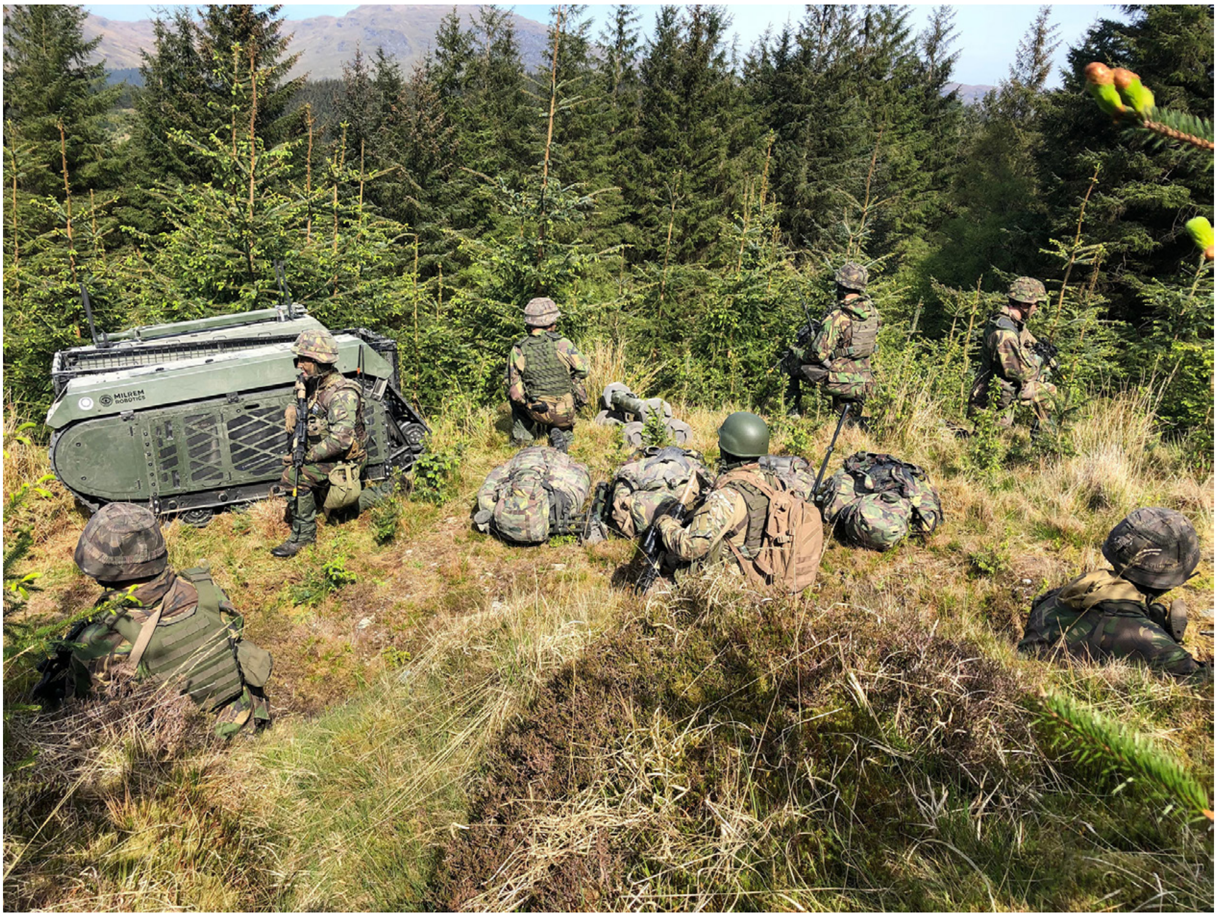


Figure 1. RNLA 13th Brigade training with THeMIS UGV in Scotland, 2019[23]

neural networks (DNN).[25] An example of this is the error rate of the visual object recognition neural network, which decreased from 25% to 3% between 2011 and 2015, compared to a human error rate of 5%.[26] While not indicative of use in complex environments, this demonstrates that AI can already supersede human abilities in certain contexts and will continue to do so in the future. The upward trend in the systems' use and their ability to carry out more functions independently, often better than their human counterparts, is causing shifts in perception of and attitudes towards RAS.[27]

[24] "Milrem Robotics Delivered Two THeMIS UGVs to the Dutch Army."

[25] Essentially, the learning process of a deep neural network is a process through which a large data set combined with high computing power makes it possible to filter through all recognizable elements of the data points in the data set in order to build a new model. Whereas older machine learning required you to first build a model to recognize the data points for what they were, with deep leaning the computer model 'teaches' itself what the defining features of all the different objects/images/etc. in the data set are. A simple example of this would be a data set of dog photos, with each photo labeled as the breed of the dog in question, from which a DNN trains to establish the features that make up a dog's breed in a way that will allow the DNN to recognize the breeds of dogs in new photos.

[26] JASON, The MITRE Corporation, "Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to DoD," 9.

[27] McLean, "Drones Are Cheap, Soldiers Are Not"; "Dilbert at War."

3. Human Agency

This chapter presents the technical complexity of RAS, discusses diverging approaches to establishing the degree of human control over RAS, and the impact this has on human-machine teaming. The chapter first presents definitions of key concepts, explains the varying degrees of autonomy and dissects the functional complexity of RAS across the life cycle, sub-system functionality, and the observe-orient-decide-act (OODA) loop. This is followed by a discussion on the determination of the acceptable degree of human control and the main factors affecting this determination, namely explainability and predictability of machines, self-learning abilities, and software updates. Building on these factors, the chapter discusses challenges in human-machine teaming that were identified in the literature review and the expert session. These are automation bias and complacency, distinction between trust or knowledge of systems and providers, interoperability issues, and the anthropomorphizing (i.e. humanizing) of the machines. While these topics are not the sole factors that elicit ethical considerations in the use of RAS, they were selected as a result of prominence in academic literature and among experts at the HCSS scenario-based expert session. The latter enabled the discussion of issues which may be underrepresented in academic literature and only arise in certain contexts.

Maintaining human agency, particularly in the context of AWS, is one of the contentious issues of debate with respect to the integration of RAS in the military domain. Human agency is a concept that encompasses "self-control, morality, memory, emotion recognition, planning, communication and thought."[28] It includes "features of self-awareness, self-consciousness and self-authorship," and as a result relates to moral agency and affects the attribution of responsibility.[29] Human control, also referred to throughout this paper as 'meaningful human control' (MHC), is an operational component of human agency, which distinguishes between human and artificial decision-making processes.[30] The term has been adopted by a number of state and non-state actors to frame the discussion on human control over and autonomy in weapon systems.[31] Although the discussion on MHC primarily concerns the operation of RAS, suggestions have been made to introduce the

[28] Gray, Gray, and Wegner, "Dimensions of Mind Perception."

[29] European Group on Ethics in Science and New Technologies (EGE), *Statement on Artificial Intelligence, Robotics and "autonomous" Systems*.

[30] MHC interrelates with "effective control", a prerequisite in public international law for legal liability and unlawful conduct. In the context of the use of RAS/AWS, the term is used alongside "effective command" to determine state responsibility. This is discussed further in Chapter 4. European Group on Ethics in Science and New Technologies (EGE); "Killer Robots and the Concept of Meaningful Human Control."

[31] MHC is not used universally, and controversy over the definition primarily centers around the degree to which there is a 'human in the loop', meaning, the degree to which a human is involved in the operating and/or decision-making process of the system. See United Nations Institute for Disarmament Research (UNIDIR, "The Weaponization of Increasingly Autonomous Technologies: Considering How Meaningful Human Control Might Move the Discussion Forward."

principle throughout the wider life cycle of a system in order to incorporate the design, procurement, testing, and decommissioning stages.[32]

3.1 Human Control

The establishment of an agreed upon definition of meaningful human control is hindered by the arbitrary nature of the acceptable degree of control and varying approaches among the nations that lead the development and application of RAS in the military domain. Beyond the challenge of defining MHC, prominent issues in establishing human control are automation bias, and system features such as self-learning abilities and software updates. While the definitions of RAS and AWS present a framework within which the systems exist, establishing human control is made difficult by the varying degrees of autonomy of systems, which resembles a spectrum rather than a clear-cut categorization. This paper defines both autonomy itself and adopts one of the commonly used classification frameworks for a system's degree of autonomy to avoid generalization of terms in the discussion.

Degrees of Autonomy

Determining the intelligence of autonomous systems is complicated by its relational nature and the human tendency to re-consider AI as computer models once a new generation of algorithms becomes achievable and operational, thus reserving the status of 'AI' for systems humans have yet to develop.[33] This affects how humans perceive AI-driven systems, particularly in the military context, and as such, which systems nation states are comfortable with rolling out within their military forces. As new systems enter general use in civil contexts, humans become accustomed to them.[34] The result is continuously shifting ethical boundaries, which dictate the acceptable use of systems and their applications. Autonomy in the context of this discussion is defined as the following:

Autonomy

Autonomy is the level of independence that humans grant a system to execute a given task. It is the condition or quality of being self-governing to achieve an assigned task based on the system's own situational awareness (integrated sensing, perceiving, analyzing), planning, and decision-making. Autonomy refers to a spectrum of automation in which independent decision-making can be tailored for a specific mission, level of risk, and degree of human-machine teaming.[35]

[32] Decommissioning is particularly relevant as the equipment can be sold to another military, whereby the risks discussed in this paper are still present, but offloaded to a third party, suggesting considerations that need to be introduced within arms control regimes.
[33] Scharre, *Army of None: Autonomous Weapons and the Future of War*, 242.
[34] Clarke, *Profiles of the Future; an Inquiry into the Limits of the Possible*.
[35] The definition is borrowed in full from Feickert et al., "U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI): Considerations for Congress."

A further distinction is drawn between automatic, automated and autonomous systems.[36] Automatic systems are simple and threshold-based, whereby their action following a sensory response is linear, immediate and highly predictable. Automated systems are more complex and consider a range of inputs and variables before acting. Autonomous systems are goal-oriented and self-directed, meaning the operator may not understand the computational process that the system used to arrive at its conclusion (see Figure 2). The distinction between automated and autonomous is difficult, as many existing RAS and AWS are forms of sophisticated automation, rather than actual autonomy.

An important difference to highlight are the two distinctive meanings of autonomy in this context. The first refers to the degree of independent intent and ability a system has to complete goals through computation not understandable by humans, often referred to as Artificial General Intelligence (AGI).[37] The second refers to the freedom of action granted to systems by humans, where a system is enabled to operate independently but is bounded by a strict set of rules, such as an automatic or automated system operating with minimal or no human oversight.[38] Most AI experts concur that AGI has not yet been achieved and is not set to be for the coming years, while complex automated systems that often operate autonomously, such as the Aegis Combat System, have been in military use for over forty years.[39] Therefore, it is crucial to distinguish between autonomy which grants machines freedom of 'thought', and autonomy which grants machines freedom of operation based on a set of rules.

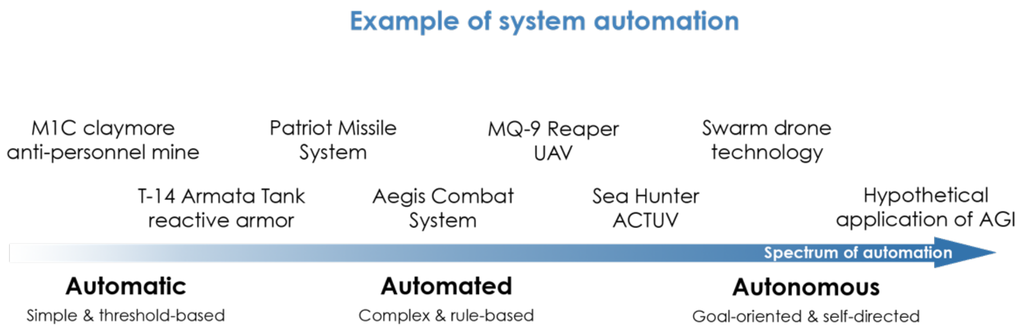


Figure 2. Examples of systems on the spectrum of automation[40]

[36] Scharre, *Army of None: Autonomous Weapons and the Future of War*, 30–31.
[37] JASON, The MITRE Corporation, "Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to DoD," 4.
[38] JASON and The MITRE Corporation, 4.
[39] Scott et al., "Modeling Artificial Intelligence and Exploring Its Impact."
[40] Adapted from Scharre, *Army of None: Autonomous Weapons and the Future of War*, 31.

Figure 2 illustrates how the degree of automation resembles a spectrum, rather than pre-defined categories, resulting primarily from technological advancements that are non-linear and manifest themselves in different machine functions. Certain sub-system functions may have a higher degree of autonomy than others, making the system more intelligent as a whole, but not enough to reach the next category, hence why the MQ-9 Reaper UAV is not considered fully within the 'autonomous' category above, for example. While the automatic–automated–autonomous scale can be perceived as a spectrum, to avoid generalizations in the discussion on autonomy, the various degrees of autonomy are placed in select categories, with the following framework adhered to throughout this paper:[41]

Direct control	A human operator has complete control over the observe-orient-decide-act (OODA) loop of the machine. This includes unmanned systems that are controlled by an operator through a machine interface. An example of this is the remote-controlled RQ-11 Raven miniature UAV.[42]
Semi-autonomous	A human operator is involved in sections of the OODA loop. An example of this are loitering munitions such as the Hero-400EC Extended Range Loitering System that requires a human to pre-identify a target but is self-guided once launched.[43]
Supervised autonomous	A human operator supervises and if necessary, intervenes in the functioning of the autonomous system, but the OODA loop functions independently as a whole. Defensive systems such as the US Aegis Combat System can effectively operate without human input after initiation, but human intervention is possible.[44]
Fully autonomous	No human is involved in the operation of the system, but most advanced militaries agree on at least a minimal requirement for a human to decide to start and shut down a system. These systems are not as widespread as those in the previous categories, but examples exist in both research and development (R&D) and actual use. A non-robotic example is the cyber defense using machine learning algorithms that learn to defend systems from new types of malicious software autonomously, in ways such as capture-the-flag games, where two intelligent systems compete to attack or defend a network or system.[45] Meanwhile, R&D of ground, naval and aerial drone swarm technology is underway in China, Russia and the US, among others, introducing the possibility of fully autonomous RAS systems that operate collectively and independently through machine-to-machine communication.[46] Since the wider-scale introduction of these systems is likely in the (near) future, we include this category to stimulate debate on the ethics of fully autonomous systems as well.

[41] "Unmanned Systems Integrated Roadmap FY 2011-2036," 46. For the purpose of this paper, the taxonomy on the degree of autonomy is adapted to the terminology used by the Netherlands Armed Forces.

[42] "RQ-11 Raven Unmanned Aerial Vehicle."

[43] "Hero-400EC Extended-Range Loitering System."

[44] "AEGIS Weapon System." It can be argued that the Aegis is a sophisticated variant of an automated system, rather than autonomous. However, due to its goal-oriented approach and multiple operating settings, it is often referred to as a supervised autonomous system. .

[45] Han et al., "Reinforcement Learning for Autonomous Defence in Software-Defined Networking."

[46] Long, "China Releases Video of 56-Boat Drone Swarm near Hong Kong"; Chung, "OFFensive Swarm-Enabled Tactics."

Perception of human control in the use of (semi-)autonomous systems is often disconnected from the actual extent of control. A prevalent heuristic is the instrumentalist perspective, through which technologies are perceived as 'tools' that are directly at the disposal, and hence, under the control of human users.[47] This is based on an underlying assumption that humans maintain agency over these tools and their operation.[48] However, the new generation of technologies is increasingly complex and purposefully designed to outperform narrow human tasks. Human agency is undermined by the cognitive inability of humans to keep up with the pace of algorithmic calculations of systems that operate with a higher degree of autonomy and independence.[49] This has different implications for operators that make decisions based on machine outputs and those that override a system, should it present incorrect or undesired outcomes. In the latter case, the ability of a human to intervene depends on the speed of the machine's operation, information available to the operator and the time delay between the human input and the system's response.[50] This illustrates that the complexity of systems undermines the instrumentalist perspective and suggests that there is a need to understand the extent to which humans maintain agency over systems and how they will continue to do so in the future. To address the complexity of systems, this paper proposes a three-part structure to identify human control in RAS. The approach establishes human control through the perspective of the system life cycle, the sub-system operational structure and the OODA loop, and is presented below.

Identifying Human Control in RAS

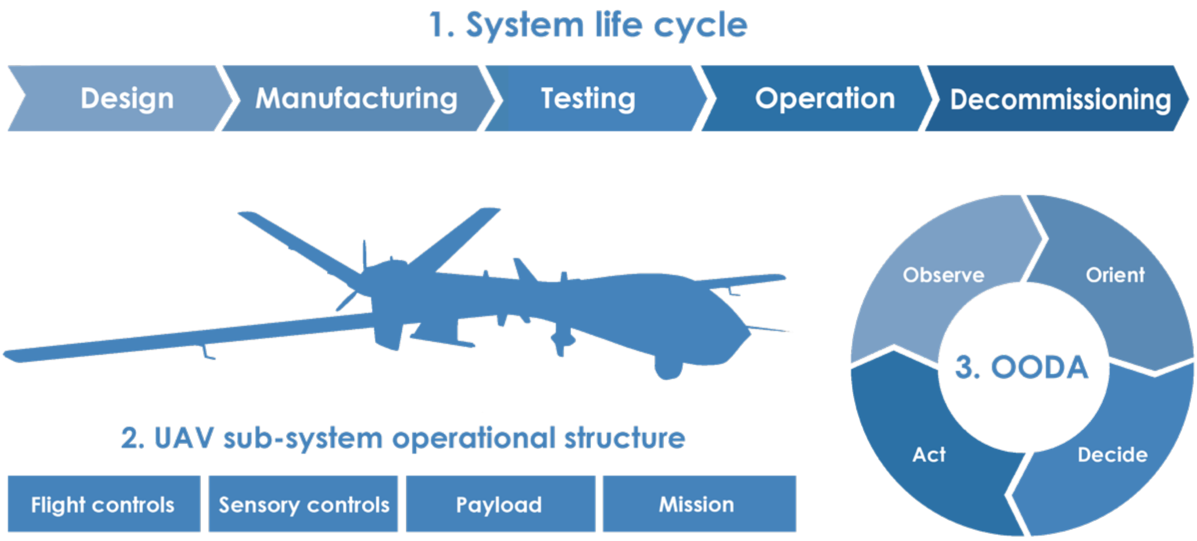


Figure 3. Various elements of human control in the operation of RAS

[47] Schwarz, "The (Im)Possibility of Meaningful Human Control for Lethal Autonomous Weapon Systems."
[48] Schwarz.
[49] Schwarz, "Intelligent Weapons Systems and Meaningful Human Control: An Uneasy Alliance," 4.
[50] Scharre, *Army of None: Autonomous Weapons and the Future of War*, 147.

Figure 3 illustrates human control through three distinct perspectives, using a UAV as an example. As the diagram above demonstrates, human control can be identified throughout the (1) life cycle of a system, meaning that an agreed upon degree of control and/or oversight is maintained throughout, or emphasized in sections such as testing and operation. A second, complimentary approach, is to identify the controversial components at the (2) sub-system level. This can separate functions such as movement control from payload in the degree of control and/or oversight required, and thus, provide a more nuanced requirement for meaningful human control in the overall operation of the system in question. A third, more detailed overview is in the decision-making cycle, presented via the (3) OODA loop. The need for the degree of human control can be identified within the specific components of the decision-making loop of a system operating in a (semi-)autonomous mode.

1. Life Cycle of RAS

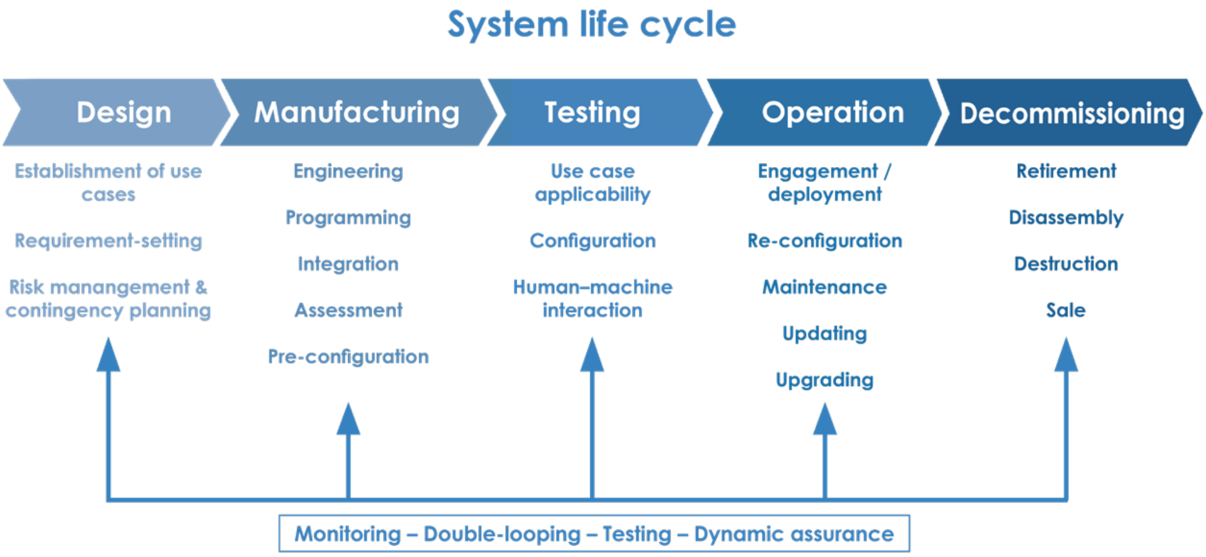


Figure 4. Life cycle perspective on human control

Design – The RAS life cycle perspective (Figure 4) seeks to contribute to the discussion on human control beyond the testing and operation sections of the system life cycle. This highlights the importance of establishing ‘ethics by design’, whereby ethical challenges are addressed early on during the design stage, thus introducing a level of control from the outset and mitigating potential shortcomings in the operation of systems ahead of time.[51] Procurement of RAS is not well-served by traditional tender processes, which highlights the need for *monitoring* system behavior, *double-looping* to previous steps, continuously *testing* and providing *dynamic assurance* throughout the

[51] Floridi et al., “From What to How - An Overview of AI Ethics Tools, Methods and Research to Translate Principles into Practices,” 14.

system's life cycle.[52] Through the involvement of end users, such as potential operators, in the *establishment of use cases* and *requirement-setting*, the appropriate user interface and user experience can be embedded into the design and manufacturing stages, which in turn can aid how operators work with, understand, and ultimately control the system. Testing and validation must be iteratively built in.[53]

Manufacturing – In both the design and manufacturing stages (which are often outsourced to private contractors) military forces need to address the risks of being dependent on external commercial actors. At these stages, inadequate supervision of contractors may result in the delivery of incomplete or low-quality equipment, exacerbating the ethical risks of their operation. To address these shortcomings, there is a need to continuously evaluate the manufacturing process and review the 'factory settings' *pre-configuration* of the system.

Testing – In testing, the people that will be working with the system in question need to gain sufficient knowledge of the system, have an intricate understanding of its function, and be able to predict the response of the system to inputs in operational environments. At this stage, limitations in human–machine teaming can be addressed through continued monitoring and double-looping of human–machine interaction. Furthermore, the *use case applicability* can be determined in the testing stage and complemented with *configuration* for the identified use cases.

Operation – During the operation stage, system deployment and use should adhere to the principles of International Humanitarian Law (IHL) and be adequately adapted for use in a specific context. Static oversight is not a guarantee of meaningful human control, which points to the requirement of system re-configuration, updating, upgrading, as well as the continuous monitoring of all implemented changes. Considering the increasing autonomy of systems, even routine activities such as maintenance require ethical considerations: it cannot be left up to chance what the output of a system may be following a tweak or an update.[54] Maintenance program escalation, which decreases the frequency of scheduled maintenance, may provide benefits such as reduced costs, but could also increase the risk of malfunction or unexpected results under certain circumstances.[55] As such, these elements need to be included within a dynamic risk assessment plan that spans the life cycle of RAS.

Decommissioning – The final stage in the life cycle is the decommissioning of RAS, both from the perspective of hardware and software longevity, as well as the accountability challenges arising from the sell-off of equipment. Certain parts will be kept internally to be reused in other RAS units, but the primary ethical challenge at the decommissioning stage is the sale or re-use of RAS

[52] Arthur van der Wees, interview, 22nd August 2019.

[53] Henderson-Sellers and Edwards, "The Object-Oriented Systems Life Cycle," 144; Lehman, "Programs, Life Cycles, and Laws of Software Evolution," 1065.

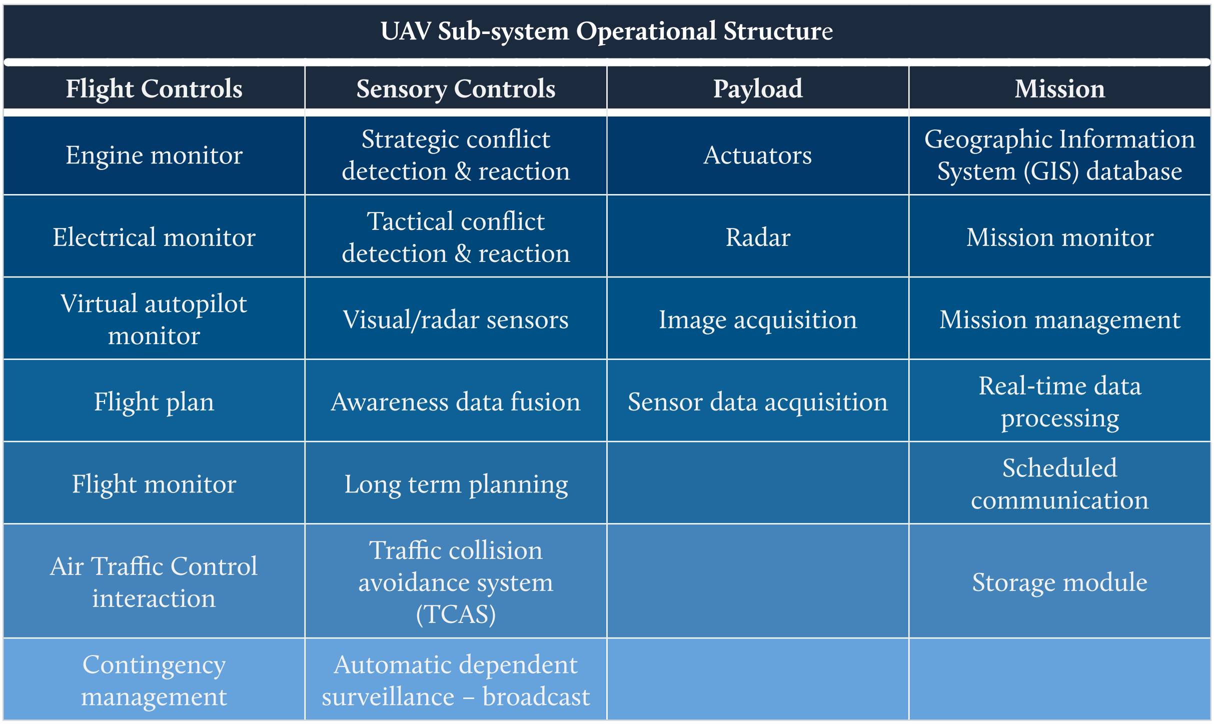
[54] European Commission, "Commission Regulation (EU) No 1321/2014 of 26 November 2014 on the Continuing Airworthiness of Aircraft and Aeronautical Products, Parts and Appliances, and on the Approval of Organisations and Personnel Involved in These Tasks."

[55] Boeing, "Maintenance Program Enhancements."

to others, be this in full or by parts. This requires the understanding of risks arising from the use of used-serviceable material.

2. Functional Complexity of RAS

Human agency should be recognized not only with respect to RAS itself, but also in regard to individual functions of RAS. Complex systems now feature a range of functions with varying degrees of autonomy. Figure 5 illustrates is the composition of a UAV, which has the following internal processes:



UAV Sub-system Operational Structure			
Flight Controls	**Sensory Controls**	**Payload**	**Mission**
Engine monitor	Strategic conflict detection & reaction	Actuators	Geographic Information System (GIS) database
Electrical monitor	Tactical conflict detection & reaction	Radar	Mission monitor
Virtual autopilot monitor	Visual/radar sensors	Image acquisition	Mission management
Flight plan	Awareness data fusion	Sensor data acquisition	Real-time data processing
Flight monitor	Long term planning		Scheduled communication
Air Traffic Control interaction	Traffic collision avoidance system (TCAS)		Storage module
Contingency management	Automatic dependent surveillance – broadcast		

Figure 5. UAV sub-system operational structure[56]

The functional complexity extends to systems beyond UAVs, where the parameters or functions may be different, but the range of sub-system autonomy may vary depending on technological advancement, the need and permissibility of use. Moreover, the number of software and hardware (sub-)components raises the questions of compatibility in the long run, whereby, elements of software may be phased out earlier than other components, rendering the remaining functions inoperable. Similarly, the software, which receives continued updates, may outlive the electronic hardware that it is based on. While throughout this paper no distinction is made between RAS as a whole and individual RAS functions, the issue is important to note when gauging the degree of autonomy of a system.

[56] Pastor et al., "An Open Architecture for the Integration of UAV Civil Applications, Aerial Vehicles."

3. OODA Loop in RAS

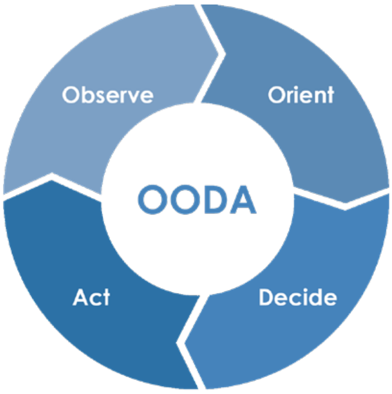


Figure 6. OODA loop

The final perspective through which this paper views human control within RAS is the observe-orient-decide-act (OODA) decision-making loop.[57] Success in a military context is derived from the ability to shorten the OODA loop more quickly than the adversary, meaning decisions are carried out at a higher pace. With the advent of automation, OODA loops have continued to shorten, in many cases shifting the role of the human from the operator to the supervisor of the system, as computational speeds exceed those of human cognition. Advanced military forces expect OODA loops to shorten to fractions of seconds in the near future, meaning that ethical principles need to be established prior to their deployment and be included in the system design.[58]

An example of the changing human role in the OODA cycle is the comparison between a semi-autonomous weapon system and a supervised autonomous system. In the case of the former, the system carries out the appropriate calculations (i.e. observes and orients) while the decision to engage is retained by the human. In supervised systems, the human may oversee the decision cycle and override the system's decision-making process, but is not involved in the system's OODA loop. While humans are still involved in the OODA loops of semi-autonomous systems, the advent of technologies such as drone swarming challenges the current construct and necessitates an understanding of how meaningful human control is retained in split-second decision-making loops.

4. Summary of the Three Perspectives

The three perspectives above present a nuanced way of establishing meaningful human control within RAS. It enables the development of guidelines based on the most challenging sections of the cycles, such as 'decide' in the OODA or the 'payload' in the sub-system structure. This enables the armed forces to continue expediting the OODA loop in sections with less ethical concern, while prioritizing the determination of human control in controversial sections. Combined, the perspectives bring to the forefront elements often disregarded in the debate on ethics, such as the design of systems and their decommissioning.

[57] Boyd, "The Essence of Winning and Losing."
[58] Scharre, *Army of None: Autonomous Weapons and the Future of War*, 23–24.

Determining Meaningful Human Control

Identifying and establishing meaningful human control can aid the process of establishing responsibility and accountability under International Humanitarian Law (IHL) in the use of AWS, particularly in the selection and engagement of targets.[59] There is a divergence in the interpretation of MHC, in terms of both the degree of control that should be required, and where and by whom this control should be maintained within the operational chain of a (semi-)autonomous system.[60] The critical issue is that there is no universally accepted definition, as on a national level each state interprets MHC to best suit their needs, while at the international level norm-setting has been impeded by states that benefit from the lack of clarity.[61] This lack of an explicit definition will continue to hamper the determination of responsibility in the use of RAS. There is an established consensus that humans are inherently responsible for actions of machines, but with increasingly complex systems that erode, or are perceived to erode human agency, it is necessary to outline exact features that would establish that the human control is 'meaningful'.[62] However, there is a further need to establish "*who* should exercise meaningful human control over *what*."[63] The current static approach of looking solely at the operator's control of the system negates the distributed nature of control that is spread across many individuals in the military decision-making cycle.[64] This reinforces the suggestion provided by this paper to view the identification of human control through the life cycle, sub-system functionality and OODA loop perspectives. Basic principles for MHC that have been proposed by normative actors, namely the ICRC and the non-governmental organization Article 36, include:

- Conscious human decisions, and timely judgment and intervention;
- Sufficient and accurate information on the outcome sought, the weapon system used and the context of its use;
- Transparency, predictability and reliability of the system linked to its design features; and
- Accountability for the functioning of the weapon system to a certain standard, such as IHL.[65]

A practical consideration is that the normative actors often "articulate an idealized version of human control divorced from the reality of warfare and the weapons that have long been considered acceptable in conducting it."[66] This reiterates the argument that (semi-)autonomous systems have been adopted by modern armed forces over four decades ago and their use has generated little controversy.[67]

[59] "Killer Robots and the Concept of Meaningful Human Control;" for more on the principles of IHL, see chapter 3.

[60] Sometimes also termed 'appropriate levels of human judgment' or sufficient human control'. See "Statement by France and Germany"; Docherty, "Heed the Call"; Sharkey, "Saying 'No!' To Lethal Autonomous Targeting."

[61] "Killer Robots Fail Key Moral, Legal Test."

[62] Santoni de Sio and van den Hoven, "Meaningful Human Control over Autonomous Systems"; Schwarz, "The (Im)Possibility of Meaningful Human Control for Lethal Autonomous Weapon Systems."the principle of "meaningful human control" has been introduced in the legal-political debate; according to this principle humans not computers and their algorithms should ultimately remain in control of, and thus morally responsible for relevant decisions about (lethal

[63] Ekelhof, "Autonomous Weapons."

[64] Ekelhof.

[65] Ekelhof.

[66] Scharre and Horowitz, "Meaningful Human Control in Weapon Systems: A Primer."

[67] Ekelhof, "Lifting the Fog of Targeting: 'Autonomous Weapons' and Human Control through the Lens of Military Targeting."

In the absence of a universal definition, it is worth considering the interpretations of MHC by key actors engaged in the deployment of RAS, namely the UK, The Netherlands, the US, Israel, China and Russia.[68] The UK, an important player in the development of autonomous systems, emphasizes that "UK weapons will always be under human control as an absolute guarantee of human oversight, authority and accountability."[69] At the same time, however, the UK has a narrower definition of RAS than most other states, defining an autonomous system as one that is

> *capable of understanding higher level intent and direction. [...] It is capable of deciding a course of action, from a number of alternatives, without depending on human oversight and control, although these may still be present. Although the overall activity of an autonomous unmanned aircraft will be predictable, individual actions may not be.*[70]

The high degree of autonomy embedded in the UK's RAS definition means that current semi-autonomous systems may be excluded from the necessity of MHC by being deemed less advanced and hence fall outside of the boundaries of RAS. Combined with the acceptance of unpredictability of certain functions of unmanned aircraft, the UK's understanding of MHC is distant to the basic principles presented by the ICRC and the non-governmental organization Article 36.[71]

The Netherlands Advisory Council on International Affairs (AIV) and Advisory Committee on Issues of Public International Law (CAVV) established that MHC is retained in the case of "an autonomous weapon [...] deployed after human consideration of aspects such as target selection, weapon selection and implementation planning, including an assessment of potential collateral damage."[72] Moreover, "in such cases, humans make informed, conscious choices regarding the use of weapons, based on adequate information about the target, the weapon in question and the context in which it is to be deployed."[73] This approach largely reflects the baselines set out by the ICRC.

Meanwhile, the US and Israel utilize the term "appropriate human judgment" rather than MHC.[74] During the 2016 Convention on Certain Conventional Weapons (CCW),[75] Israel argued that appropriate human judgment is already "built into the development of weapons systems, including at the design, testing, and deployment phases, and thus requiring meaningful human control is unnecessary."[76] The US also actively considers the entire RAS life cycle, meaning that "systems will go through rigorous hardware and software verification and validation (V&V) and realistic

[68] Most are derived from country statements at the Convention on Certain Conventional Weapons (CCW) meetings on lethal autonomous weapon systems.

[69] Evans, "Too Early for a Ban: The U.S. and U.K. Positions on Lethal Autonomous Weapons Systems."

[70] Development, Concepts and Doctrine Centre, "Unmanned Aircraft Systems - Joint Doctrine Publication 0-30.2."

[71] The UK's definition of RAS has 'higher level' expectations, which are beyond the majority of systems operated at this time. As a result, the discussion on MHC for UK's understanding of RAS may only apply to the complex systems envisioned with a higher degree of autonomy and thus downplay the degree of independence of existing systems.

[72] Netherlands Advisory Council on International Affairs, "Autonomous Weapon Systems: The Need for Meaningful Human Control."

[73] Netherlands Advisory Council on International Affairs.

[74] "Killer Robots Fail Key Moral, Legal Test."

[75] Lewis, "AI and Autonomy in War: Understanding and Mitigating Risks"; "Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects."

[76] "Killer Robots Fail Key Moral, Legal Test."

system developmental and operational test and evaluation (T&E) [...]."[77] The US Department of Defense directive on autonomy in weapon systems requires "traceable feedback on system status", explainability and predictability features, and has safety considerations for the human–machine interface.[78] It is therefore evident that countries have strongly varying positions on MHC, but some leading actors in RAS development agree on the importance of establishing some degree of human control in the design and procurement processes, rather than just in training and operations.

At the 2016 CCW meeting, China stated that "[t]he mode of human involvement and the human role [...] requires a strict definition and cannot be replaced by such vague concepts as 'human judgment' or 'meaningful human control'."[79] Internationally, China maintains its position that controllability remains a priority for any (semi-)autonomous military technology. However, on the same day that China reiterated its support for the development of a binding protocol banning the use of fully autonomous weapons at the 2018 CCW meeting, the country's air force published a statement that clearly demonstrated China intends to develop such systems anyway.[80] China is likely to advocate for international agreements that leave its own view on MHC slightly ambiguous, while delegitimizing the moral position of actors such as the US that have opposed adopting new laws on this topic just yet.[81]

Despite the official statements at the CCW meetings,[82] the Russian defense ministry is clear about its intention to develop autonomous weapon systems, with some arguing that "Russia seeks to completely automate the battlefield."[83] Similarly to China, the country is on track to develop swarm units, which are groupings of autonomous systems that are inherently difficult to maintain human control over once deployed. The announcement by weapons manufacturer Kalashnikov that it will develop "a series of autonomous weapons using neural networks trained to autonomously track targets and fire on them" and Degtyarev's development of the "suicide tank" match defense officials' enthusiasm for robotization and lack of interest in human control as a prerequisite—no matter the definition.[84]

[77] US Department of Defense, "Directive 3009.09."
[78] US Department of Defense.
[79] "The Position Paper Submitted by the Chinese Delegation to CCW 5th Review Conference."
[80] Kania, "China's Strategic Ambiguity and Shifting Approach to Lethal Autonomous Weapons Systems," April 17, 2018; Klare, "Autonomous Weapons Systems and the Laws of War"; Mohanty, "Lethal Autonomous Dragon."
[81] Kania, "China's Strategic Ambiguity and Shifting Approach to Lethal Autonomous Weapons Systems," April 17, 2018; Kania, "Battlefield Singularity."
[82] "Statement of the Head of the Russian Federation Delegation, Director of the Department for Nonproliferation and Arms Control of the Russian Ministry for Foreign Affairs V.Yermakov at the Meeting of the State-Parties of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons on Item 7 of the Agenda 'General Exchange of Views', Geneva, November 21, 2018."
[83] Sharkey, "Killer Robots From Russia Without Love."
[84] As for the suicide tank, "Once launched it can navigate autonomously to a target in silent mode and then explode with a powerful force to destroy other tanks or entire buildings.". See Sharkey; Gilbert, "Russian Weapons Maker Kalashnikov Developing Killer AI Robots."

Traceability, Explainability and Predictability

A fundamental aspect of maintaining MHC is the operator's understanding of the algorithmic process' parameters, the outcomes presented as a result of the computation, and the ability to explain the machine's path to conclusion after the fact. Explainability is a prerequisite to determining some degree of operator's responsibility for the actions of a system. One of the ethical criticisms of RAS, and AI in particular, is the current lack of algorithmic transparency. Algorithms such as neural networks suffer from opacity as they operate as 'black boxes', whereby the path taken by the algorithm to arrive at the conclusion is often not traceable.[85] Beyond the 'black box' effect, algorithmic opacity arises from technical illiteracy, whereby creators write poorly structured code, or programmers on the receiving end are unable to understand the creator's intent. Opacity is reinforced by the complexity, self-learning capabilities and scale of algorithms, with systems such as the F-35 fighter jet and self-driving vehicles requiring 24 million and 100 million lines of code, respectively.[86] As a result, algorithmic activity may not be traced and hence, in the event of a malfunction, the cause of the failure will not be determined rapidly. The diminished understanding an operator has of such systems reduces their ability to predict and/or explain the system's reasoning process. This may undermine the control that the operator has over the outcomes and hence, the responsibility for its (mis)use.[87] However, given the sheer scale of code in complex systems, the expectation for the operator to understand granular functions, especially in the operation stage of the life cycle is unrealistic. Progress is being made in understanding the internal workings of algorithms, by means such as the Local Interpretable Model-Agnostic Explanations, an algorithm that explains the prediction of any classifier algorithm in a method interpretable by humans.[88] An alternative is the use of 'logic flow diagrams', which summarize sets of code and enable the operator or supervisor to trace the process via critical junctures such as the OODA loop steps, thus maintaining a macro perspective of the system performance.[89] This approach reflects the use of a vehicle without explicit understanding of its mechanical functions but with knowledge of the error signals displayed on the driver dashboard and their meaning.

Algorithmic systems predominantly operate based on historical training data to make future assessments and predictions. The need for quantification renders contextual non-numerical data—such as an individual's behavior and body language that are observed rather than measured—potentially invalid in the algorithmic decision-making process. This means elements that cannot be easily quantified are likely to be excluded from the calculation.[90] The result is a system that can

[85] Preece, "Asking 'Why' in AI: Explainability of Intelligent Systems – Perspectives and Challenges"; Matthias, "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata," 178–79.

[86] Burrell, "How the Machine 'Thinks'"; Mittelstadt et al., "The Ethics of Algorithms," 3–7; Scharre, *Army of None: Autonomous Weapons and the Future of War*, 157.

[87] Preece, "Asking 'Why' in AI: Explainability of Intelligent Systems – Perspectives and Challenges"; Mittelstadt et al., "The Ethics of Algorithms," 5, 10–12.

[88] Ribeiro, Singh, and Guestrin, ""Why Should I Trust You?"

[89] Wu et al., "Research and Application of Code Automatic Generation Algorithm Based on Structured Flowchart"; Kumar et al., "Algorithms, Flowcharts, Data Types and Pseudocode."

[90] Schwarz, "The (Im)Possibility of Meaningful Human Control for Lethal Autonomous Weapon Systems."

function successfully in a controlled environment with defined parameters, but in a real-world scenario, where parameters are less defined, algorithms present outcomes based on a biased set of numerical inputs.[91] Moreover, the algorithms are often "embedded at the backend of systems, [...] with no consumer-facing interface. Their operations are mainly unknown, unseen, and with impacts that take enormous effort to detect."[92] In a high-intensity setting, this further establishes the need for extensive explainability, as the operator has to be acquainted with the system's intricate parameters and be familiar with its shortcomings between controlled and real-world scenarios.

When combined with other risks discussed in this chapter, the 'black box' effect can impede the functioning of human–machine teaming. Beyond understanding the reasoning process the system undertook to arrive at the conclusion, it is important for an operator to be able to anticipate how RAS will react in any given situation, particularly when it comes to a real-world scenario following controlled testing. A system may be predictable in a controlled environment for a programmer but may not maintain the same properties in a real-world scenario for an operator, supervisor and/or commander. Moreover, predictability of actions does not guarantee predictability of outcomes, which once again, can be influenced by the operational environment.[93] The discrepancy between training and combat, which may result in a predictability gap, is discussed further in Chapter 3.2.

Self-learning Abilities and Software Updates

The evolutionary nature of algorithm-driven systems, both as a result of self-learning properties and software updates, has the potential to considerably affect explainability of systems' actions. Self-learning AI that independently develops its understanding of the surrounding environment may limit human control over the system's operation. This is underpinned by the exponential growth of the capabilities of machine learning (ML) algorithms, such as, among others, neural networks and reinforcement learning.[94] As systems become increasingly complex, "in a steady progression the programmer role changes from coder to creator of software organisms."[95] Programmers are transitioning from maintaining control over the software code, to setting the algorithmic parameters and, depending on the algorithm, the network architecture for the algorithm to operate within. Reinforcement learning algorithms are particularly challenging, as they are designed to learn from their immediate environment.[96] The result of this is that next-generation algorithms no longer operate on pre-determined rules and can change their functionality, meaning humans often cannot understand the calculation made to arrive at the conclusion.

[91] Schwarz, "Intelligent Weapons Systems and Meaningful Human Control: An Uneasy Alliance," 11.
[92] Crawford and Whittaker, "Artificial Intelligence Is Hard to See."
[93] United Nations Institute for Disarmament Research (UNIDIR, "The Weaponization of Increasingly Autonomous Technologies: Considering How Meaningful Human Control Might Move the Discussion Forward."
[94] Scott et al., "Modeling Artificial Intelligence and Exploring Its Impact."
[95] Matthias, "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata."
[96] An example of this is the AlphaGo Zero algorithm that mastered the game Go without making use of historical training datasets based on human inputs. United Nations Institute for Disarmament Research (UNIDIR, "The Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence."

The evolving functionality is compounded by the involvement of third parties, often private companies, which are responsible for designing and supplying the system. This illustrates how human control is affected in the aforementioned life cycle and OODA perspectives. A further layer of complexity is added by the distancing of the programmer from the system, meaning the response time to faults and malfunctions is increased.[97] While the private contractors who design and manufacture the systems are now often deployed alongside the military, the potential inability of their military counterparts to understand RAS undermines meaningful control in the system's use, as it is the military operator who makes substantive decisions over the system's utilization. Depending on one's interpretation of MHC, there is potential for a loss of human control over time. This is a result of not only increasing sophistication in the AI that powers the RAS, but also the subsequent software patches that may reduce the military operators' understanding of the system over time.

This issue arose in at least two of the four scenarios in the HCSS expert session. Participants argued that operators would require training with the machine following each subsequent software update, as it could alter the machine's behavior and, as a result, the predictability of its outputs.[98] In reference to scenario 3, one participant noted the difficulty brought about by software updates, which meant that personnel had to get used to the change in the machine's method of operation, but the update could also unwittingly alter other capabilities. This begged the question of how much training is reasonable or necessary under these circumstances to ensure sufficient (re)familiarization with the updated system. Therefore, the RNLA would need to determine whether retraining is necessary for software updates involving all components of RAS or only for specific, pre-defined components.

To highlight the compounding effect of the aforementioned factors, an interesting example from civil aviation is that of the two Boeing 737 MAX 8 crashes, and the subsequent grounding of all MAX 8 aircraft worldwide in the Spring of 2019.[99] While the case is outside the military domain, it involves Boeing, a manufacturer which doubles as a defense contractor and highlights the challenges of relying on external contractors, as well as continuous software updating and system modification. In a cost-cutting bid, Boeing supplemented major mechanical engine re-design in the 737 MAX 8 with sensors and an additional automated system. The system aimed

[97] Matthias, "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata."

[98] In scenario 2 (testbed), experts argued that the system would have to undergo testing after each software update, to ensure that the operator has up-to-date understanding of the system's behavior.

[99] Based on preliminary findings, the cause of the two crashes (Lion Air Flight 610 & Ethiopian Airlines Flight 302) involving the Boeing 737 MAX 8 has been a malfunction of the Maneuvering Characteristics Augmentation System (MCAS). The system was introduced to offset the engine design changes made to the 737 MAX 8 from the previous models in the 737 series. More specifically, the change was the relocation of the engine position under the wings, resulting in a shift of the centre of gravity of the aircraft. To control for this, Boeing installed the MCAS, which was supposed to use sensors to indicate to a computer if the aircraft was stalling mid-air. However, the system also reacted in cases where the sensor input was contradictory. In both incidents, the system falsely identified the aircraft's angle-of-attack as excessively high and sent the aircraft into a nose dive as to prevent it from stalling midair. The aircraft were not in fact stalling, and the MCAS misidentifications and automatic reaction, combined with the planes' close proximity to the ground (as both incidents occurred at take-off) left the pilots no time to override the MCAS and take manual control of the aircraft. Travis, "How the Boeing 737 Max Disaster Looks to a Software Developer"; Lu et al., "From 8,600 Flights to Zero: Grounding the Boeing 737 Max 8."

to compensate for the change in engine design and its new position under the wing.[100] The pilot manual included with the new aircraft did not sufficiently inform pilots familiar with other Boeing 737 variants of the changes introduced in the MAX 8.[101] This meant the pilots were not well trained with the machine interface, and as has been suggested following the incidents, led to the inability of pilots to override the system in a short span of time when it malfunctioned. The operation of a highly automated system—the 737 MAX 8—was hampered by substantial functionality alterations that were not well communicated to its users. Boeing presented the aircraft as closely related to previous 737 models, thus suggesting that extensive re-training was not necessary, in turn resulting in incomplete preparation of the pilots to use the new aircraft. The delivery of the unsafe 737 MAX 8's highlights the additional risk of poor oversight of private contractors, an issue discussed later in this chapter.[102]

The legal cases following the incidents are on-going, but the issue of fragmented responsibility is already evident. Who is responsible for the incidents? At the macro level, is it the airline companies, Boeing, pilot training organizations and flight simulator operators, regulators of airlines in the host countries or the regulators of Boeing's host country (in this case the US Federal Aviation Authority (FAA))? At the micro level, is it the pilots, airline engineers, Boeing engineers, Boeing software developers accountable or the FAA inspectors? The issue of attributing responsibility is further complicated by public–private and cross-border divides, whereby accountability can be hampered by national laws and (international) contracts. While this case is outside the military context, it clearly demonstrates the discussed issues unfolding in a real-life context and reflects the issue of distributed responsibility in the military decision-making cycle, which also involves many actors, from programmers and manufacturers, to operators and commanders.

3.2 Human–Machine Teaming

The relationship between the military personnel and operators, and the machines they work with, is another important aspect of human agency. How RAS are developed and deployed will deeply impact the people working with them.[103] Aside from determining the level of MHC and being able to understand and explain the reasoning process of the system, human–machine interaction encompasses several other issues, raised both in the literature and the expert session. These are automation bias, interoperability challenges, trust in the manufacturer of the system versus knowledge of the system itself, comparability of testing environments to real scenarios and the anthropomorphizing of machines.

[100] Travis, "How the Boeing 737 Max Disaster Looks to a Software Developer."
[101] Hawkins, "Everything You Need to Know about the Boeing 737 Max Airplane Crashes."
[102] Stewart, "The Boeing 737 Max 8 Crashes and Controversy, Explained."
[103] Roff and Danks, "'Trust but Verify.'"

Automation Bias and Complacency

Human overreliance on and uncritical trust in computer-based decision-making, otherwise known as 'automation bias', impacts the level of control that operators can exercise over RAS.[104] This is a human tendency to ignore or not seek out contradictory information to the outputs of an automated process, due to a perception of machines' superiority in accuracy.[105] 'Automation complacency' is not dissimilar, but while automation bias refers to excessive trust in a system, automation complacency concerns substandard attention to and monitoring of a system's output, on the assumption that the output is reliable.[106] Both bias and complacency lead to problems of process malfunction misidentification, anomalies and failure, as well as delays in the response time of human intervention resulting from insufficient oversight. The latter is no less critical than outright failures in oversight, as in fast-paced combat situations, an untimely response can have serious implications for the outcomes of the use of RAS and AWS.[107]

As a result of automation bias, meaningful human control is reduced. Humans may place a disproportionate amount of trust in the automated processes they are meant to control or supervise. An example of this is the shooting down of an Iranian passenger jet in the Persian Gulf by the United States Ship (USS) Vincennes in 1988.[108] Amidst an engagement between the USS Vincennes and Iranian forces, the automatic targeting-and-firing system Aegis misinterpreted the passenger jet for a military fighter jet and the naval vessel crew shot down the civilian aircraft. This highlights a failure to recognize and challenge shortcomings in a computerized decision-making process, particularly in a high-intensity combat situation when what appears to be an imminent threat can lead to a lethal counterattack. While the error occurred due to an incorrectly pre-selected operation setting, the automation of the weapon system in this instance significantly reduced the time for human decision-making and intervention. The operators' lack of due diligence in this case highlights the risks of the coupling of narrowing response time frames with automation bias and/or automation complacency.

Consideration of automation bias and complacency is therefore critical in establishing meaningful human control in the use of RAS, as "technologists tend to push to automate tasks as fully as possible."[109] There is a need for explicit understanding of how increased process automation affects human cognition, and in turn human–machine teaming.[110] Determining the balance between the benefits of automation and its risks is particularly important in a military context, where bias and complacency can reduce explainability and, in turn, the responsibility of operators. The RNLA

[104] Cummings, "Automation Bias in Intelligent Time Critical Decision Support Systems."
[105] Cummings.
[106] Parasuraman and Manzey, "Complacency and Bias in Human Use of Automation: An Attentional Integration," 382.
[107] NASA's Aviation Safety Reporting System (ASRS) defines complacency as "the state of self- satisfaction that is often coupled with unawareness of impending trouble," see Bhana, "By the Book - Good Written Guidance and Procedures Reduce Pilots' Automation Complacency"; Parasuraman and Manzey, "Complacency and Bias in Human Use of Automation: An Attentional Integration."
[108] Kania, "The Critical Human Element in the Machine Age of Warfare."
[109] Miller and Parasuraman, "Designing for Flexible Interaction Between Humans and Automation," 58.
[110] Hoijtink and Leese, *Technology and Agency in International Relations*, 50–53.

should seek to understand at what degree of autonomy the limitations of human cognition in the oversight of RAS offset the benefit of increased autonomy. An alternative function is one similar to operator assist, whereby the systems enhance human functions, rather than replacing them entirely.[111] In this instance a critical consideration is testing and validation of human–machine interfaces that diminish the effects of automation bias and complacency.

Trust or Knowledge

Another point of discussion is that RAS—whether as logistics, weapons, or otherwise—tend to be developed in order to enhance militaries' all-round capabilities.[112] This justification assumes both that the RAS will function as intended, and that the users and/or operators trust the systems to the extent that these can function as intended.[113] This issue is particularly potent when considering the role of third-party contractors in supplying RAS, particularly AWS. There is a risk of the military outsourcing excessive control to the contractor and not being fully informed on the reasoning process of RAS, particularly when contractors are involved in the operation of the system. This relates to the case involving the Boeing 737 MAX 8, which supplied hundreds of aircraft to over 40 airlines worldwide before the fatal issue was identified.[114] This requires a due-diligence process before and throughout the engagement with external contractors. Oversight has to be maintained in the design of the systems and throughout their deployment, since as long as they continue to function, they will depend on externally developed software and hardware to do so. Therefore, the RNLA need to ensure that it has thorough knowledge of the functional parameters of the (semi-) autonomous systems they purchase, rather than basing the acquisition and use on the trust placed in the contractor and the operator of the system.

Interoperability with Partner Forces

A variation of the 'trust or knowledge' issue was identified during the expert session. It concerns how the military can engage in combat alongside technologically advanced allied military forces operating RAS that the former has not yet worked with.[115] In the expert session discussions, participants concluded that for the joint use of an autonomous system in the military, it is crucial that the troops using it fully understand the system they are using and can predict the behavior of RAS in the situation or environment that the system is used in. However, the necessary level of (training) experience with a system, or how much prior information on the system from a partner, remained a point of discussion, and is clearly a topic that requires further research.

[111] JASON and The MITRE Corporation, "Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to DoD," 54.
[112] Roff and Danks, "'Trust but Verify.'"
[113] Development, Concepts and Doctrine Centre, *Human-Machine Teaming*.
[114] Lu et al., "From 8,600 Flights to Zero: Grounding the Boeing 737 Max 8."
[115] In scenario 3, Dutch soldiers, hypothetically, are faced with a choice between deploying tired Dutch soldiers to secure a facility, or deploying Danish-made and -operated LAWS, the exact parameters and reasoning process of which are unknown to the Dutch contingent.

Opinions on the level of training ranged from troops needing to be well acquainted with and have trained with the system for years on the one hand, to troops in collaboration with trusted partners receiving a certain amount of information on the systems used, including testing reports, relevant indicators and embedded rules of engagement (ROE). Difference in levels of not only confidence, but also trust in fellow soldiers as opposed to RAS, partly comes down to the extent to which one can understand the reasoning process of each. Even if a human acts irrationally, there is a certain reasoning process behind the actions that can be explained and likely understood afterward. Therefore, a level of understanding of the reasoning process that guides the actions of RAS is needed and this should likely be communicated to partner forces that are collectively engaged in a single operational theatre. This issue highlights the need for standardization frameworks between allied forces that deploy and operate RAS during joint missions. This can further be extended to incorporate interoperability with private military contractors that are deployed alongside many advanced militaries in operational theatres today.

Testing Environments

Most RAS are tested in safe and controlled environments, so the response of the systems and the operators in a real, high-intensity and uncontrolled scenario is often unknown. For the operator, it is important to determine how they will respond to the machine's outputs, particularly when the intervention or decision-making timeframe is narrowed by the rapid computational processes. An example of where this becomes problematic is the 2004 friendly fire incident involving the downing of two British and US fighter jets by a US Patriot system over Iraq, resulting in the deaths of three aircrew.[116] The lack of training of the operators and their unfamiliarity with the machine interface resulted in the inability of the operators to intervene in the Patriot's decision to fire at allied aircraft.[117] As a result, the lack of training with the system and insufficient familiarity with the reasoning process of the Patriot in this case, further reinforced the operator's dependency on the conclusions of the system's internal processes.

From the perspective of RAS, it is unclear how a system, particularly with self-learning abilities, will respond to situational uncertainty and nuance otherwise not present in controlled testing environments, particularly if the system depends on machine learning algorithms. The issue was discussed in reaction to scenario 1 (in Appendix B) in the expert session, where experts doubted the ability of the AWS to distinguish its high-value target from other militants and civilians in a poorly lit cave system. There is an evident need to simulate highly realistic combat scenarios and even test equipment outside of controlled environments. However controversial, Russia has used its recent Syria campaign to test various autonomous systems in combat, namely an underwater unmanned system and an electronic warfare unit.[118] The RNLA may therefore consider studying the testing approaches of other states and determine how they can best emulate it without violating ethical principles.

[116] Cummings, "Automation Bias in Intelligent Time Critical Decision Support Systems."
[117] Cummings.
[118] Grishenko, "Российский Подводный Робот Выполнил Боевую Задачу в Сирии"; Bendett, "In AI, Russia Is Hustling to Catch Up."

Anthropomorphizing the machine

The final point of discussion is that individuals interacting with artificial systems tend to anthropomorphize (humanize) them, whereby they "attribute minds to computers and perceive robots as agents."[119] When RAS are "seen as more than just a tool to achieve an effect," this may hinder the intended functions.[120] This highlights the human need to attribute features to inanimate objects and paradoxically relates to the instrumentalist perspective introduced earlier in Chapter 3.1.[121] The "efforts at building self-explicating machines in their more sophisticated forms now adopt the metaphor of the machine as an expert and the user as a novice or student,"[122] demonstrating that humans are transitioning from perceiving machines as tools under their control (instrumentalist perspective), to humanizing them as they increasingly replicate human functions. This is a result of design that deliberately seeks to develop systems that exceed our control and cognitive abilities.[123] The limitations that result from humanizing machines are primarily manifested in two ways. The first is the operator becoming attached to the machine, which defeats the purpose of having RAS replace soldiers in combat and influences the operator's risk perception of the situation.[124] The second is the operator humanizing the machine and anticipating a human way of thinking, thus overseeing the limitations of algorithmic outputs.[125] The risk is that the operator is unable to develop a mental model to cope with the system and handle the system's failures, resulting in undesired effects within human–machine teaming.[126]

3.3 Summary

This chapter has highlighted the difficulty in identifying the extent of autonomy as well as determining meaningful human control and establishing it in practice. While ethicists debate practitioners on whether autonomous weapons should be banned, less controversial applications of RAS will continue to permeate the military domain. Establishing meaningful human control should be preceded by the identification of the type of autonomy displayed in a system, within which sub-system functions its present, and how this affects the OODA loop of RAS. As the armed forces seek to harness efficiency gains of AI and maintain their competitive edge, research and deployment of RAS is likely to persevere.

[119] Verdiesen, "Agency Perception and Moral Values Related to Autonomous Weapons," 96; Schwarz, "Intelligent Weapons Systems and Meaningful Human Control: An Uneasy Alliance," 4,11.

[120] Krishman, *Killer Robots: Legality and Ethicality of Autonomous Weapons*; Hsu, "Real Soldiers Love Their Robot Brethren"; Schwarz, "The (Im)Possibility of Meaningful Human Control for Lethal Autonomous Weapon Systems"; Robert, "The Growing Problem of Humanizing Robots."

[121] Sharkey, "The Evitability of Autonomous Robot Warfare."

[122] Suchman, *Human-Machine Reconfigurations: Plans and Situated Actions*.

[123] Gunkel, "Other Things: AI, Robots and Society," 60; Schwarz, "Intelligent Weapons Systems and Meaningful Human Control: An Uneasy Alliance," 4, 11.

[124] Giger et al., "Humanization of Robots."

[125] Robert, "The Growing Problem of Humanizing Robots."

[126] Brenton and Bosse, "The Cognitive Costs and Benefits of Automation."

With countries seeking to shorten the OODA loop, there is an opportunity in distinguishing between ethically controversial and undisputed functions of RAS and using this to streamline automation. The former, primarily concerning elements of weapons delivery, require further consideration and addressing of ethical questions, while the latter, such as movement or sensory control, can continue to be automated. To confidently pronounce that an operator/supervisor has meaningful control at the 'decide' stage of the OODA loop requires the assertion that human control was maintained in the design and testing of the system as well as within its sub-system functions. This, for example, involves understanding of sensory outputs that inform the 'observe' and 'orient' sections of the OODA loop, before a decision can be made. Following this, establishing MHC requires an understanding of where, how and by whom it should be maintained. Finally, challenges around maintaining MHC, among which are cognitive limitations in human–machine interaction, shortcomings in machine development, and risks of procurement from third parties, must be addressed to ensure that responsibility for the deployment and the use of RAS can be established, and that those responsible can be held accountable.

4. Human Dignity

International Humanitarian Law (IHL) has a strong footing in ethics, and its basis lies in the attempt to attain and maintain respect for the dignity of human life in the disarray of war. To be exact, IHL refers to international rules intended to protect people and property that are, or may be, affected by international or non-international conflict through setting limits on how conflicting parties may choose their methods and means of warfare.[127] Presently, there are no specific bans or regulations under international law that make RAS unlawful per se. At the same time, however, there can be no doubt that their use in conflict still means that those deploying the systems are bound by IHL, creating obligations for human combatants to ensure lawful use of RAS, and autonomous weapon systems in particular.[128] IHL principles are therefore an important part of addressing ethical issues pertaining to military RAS in general, and the use of force when deploying AWS.

Currently there are no offensive (supervised) AWS ready for deployment that in offensive situations could satisfy IHL obligations.[129] This makes for a relatively easy conclusion on non-deployment at the present stage.[130] However, it was highlighted in the expert discussions that once the technical ability to comply with said IHL obligations is available, the difficulty will be assessing under what circumstances the use of AWS could be permissible, and what we as a society deem crucial to maintaining some level of humanity on the battlefield.

This chapter on human dignity is split into two sub-chapters, aimed at addressing the ethical discussions outlined in the previous two paragraphs. In Chapter 4.1 the key principles that govern hostilities are laid out and structured along the lines of International Humanitarian Law.[131] This sub-chapter also addresses the ways in which these principles may affect or be affected by increased integration of RAS into the armed forces. The next sub-chapter, 4.2, describes the main arguments that feature in debates on how increased use of military RAS may affect the status of and respect for human dignity. Two main points of contention were identified before the expert session for the topic of human dignity: whether decisions that have always been inherently human could and should be substituted with computer processes—especially if they involve life-and-death situations; and whether there may be a point in time, or a particular situation, where substituting certain

[127] Bouvier, "International Humanitarian Law and the Law of Armed Conflict," 13.
[128] Davison, "Autonomous Weapon Systems."
[129] Boothby, *New Technologies and the Law in War and Peace*; Chehtman, "New Technologies Symposium."
[130] Boothby, *New Technologies and the Law in War and Peace*.
[131] In military spheres preference is sometimes given to the phrase 'law of armed conflict' (LOAC) rather than International Humanitarian Law. However, the authors will refer throughout this paper to IHL as it is more widely used. See e.g. Bouvier, "International Humanitarian Law and the Law of Armed Conflict," 13.

human tasks or operations with RAS may be considered more ethical, rather than less. These are two central questions that feature in the discussion of how RAS may enhance or hamper respect for human dignity.

4.1 Ethics and International Humanitarian Law

As stated above, this section lays out the key principles of IHL applicable in armed conflict. These are proportionality, military necessity, distinction, and, as a more general guiding principle underpinning the conception of ethics, humanity.[132] Throughout each sub-section there will be an explanation of what each principle entails in general, as well as how it affects or is affected by RAS—and by AWS in particular.

Proportionality

The first principle of IHL dictates that actions should always be proportionate. 'Proportionate' in this context means that expected incidental harm to civilians or civilian objects—also known as 'collateral damage'—should not be excessive in relation to the concrete and direct military advantage anticipated.[133] The standard by which this is assessed is that of a "reasonable commander or combatant who weighs the expected collateral damage against the anticipated military advantage in good faith, based on information available at the time of the attack."[134] Whether this standard suffices when it comes to using RAS or how exactly it would apply, makes the breakdown of RAS and human control illustrated in Chapter 3.1 useful as a guideline.

The degree of leeway offered by the proportionality principle as described in IHL has often been interpreted differently, and the interpretation has also shifted over time. Proportionality is viewed both as a permissive and a restrictive principle. On the one hand, the fact that states are themselves responsible for weighing military advantage from certain actions against the degree of civilian damage, could be viewed as permissive. On the other hand, the principle could be restrictive in that it may hamper military objectives, as more scrutiny is aimed at the justification behind individual attacks. With a shift from interstate conflict to more asymmetric forms of warfare, there tends not to be a clear-cut start or finish to a conflict, and properly establishing 'military advantage' can be difficult even for experienced military commanders. Therefore, one can imagine the challenge of ensuring clear compliance with the proportionality principle by an autonomous system, be it defensive of offensive.

[132] Within LOAC the principle of humanity, also referred to as the Martens Clause, is often discussed in more straightforward terms as meaning the prevention of unnecessary suffering.

[133] Additional Protocol I, Article 51(5)(b) concerning the conduct of hostilities prohibits attacks when the civilian harm would be "excessive in relation to the concrete and direct military advantage anticipated." See Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I).

[134] Netherlands Advisory Council on International Affairs, "Autonomous Weapon Systems: The Need for Meaningful Human Control," 24.

For AWS in particular, a difference can be made between the so-called 'easy proportionality problem' and 'hard proportionality problem'. The former concerns how to minimize collateral damage by using the most appropriate weapon and target, or in other words, taking all necessary precautions to minimize the damage done to civilians and civilian objects.[135] The 'hard proportionality problem' concerns the decision on whether or not to use force in the first place. This decision depends on how a commander weighs the balance between civilian lives and the wider military goal of the mission.[136] The 'hard' problem therefore concerns contextual factors beyond the specific situation at hand. Machines may be(come) better than humans at quickly assessing quantitative, computable elements of an attack, such as blast effect or number of civilian casualties. However, qualitative elements like the direct and indirect military advantage versus the civilian damage done remains, and may continue to remain for the foreseeable future, in better hands with humans.[137] The appreciation and weighing of a certain attack within the complicated context of a mission's larger military strategic aims, as opposed to within only the attack itself, involves difficult and occasionally morality-heavy decisions for commanders, and remains therefore a point of concern for the deployment of RAS.[138]

It appears that people hold RAS to higher standards than humans when it comes to accepting mistakes on the battlefield. In one of the expert session scenarios, an AWS had an accuracy rate of 99.95%, but the 0.05% was already viewed as a major issue in light of civilians present in the operational environment.[139] Most participants were undecided in this particular scenario. On the one hand, there were major concerns over the machine's ability to fulfill the requirements of proportionality. It was also acknowledged, however, that the use of RAS in the scenario would—in the case that no mistakes occurred—be the most effective and least risky option for the troops, given the almost impossible circumstances of the scenario.

Military Necessity

The second principle of IHL, albeit related to proportionality, is the principle of military necessity. Following this principle, the use of RAS may only result in the use of force for legitimate military objectives, and every injury caused—even against enemy combatants—is only excusable in as far as it was absolutely necessary.

[135] Ekelhof, "Autonome Wapensystemen: Wat We Moeten Weten over de Toepassing van Het Humanitair Oorlogsrecht En de Menselijke Rol in Militaire Besluitvorming," 198.
[136] See
[137] Ekelhof, "Autonome Wapensystemen: Wat We Moeten Weten over de Toepassing van Het Humanitair Oorlogsrecht En de Menselijke Rol in Militaire Besluitvorming," 198.
[138] van den Boogaard, "Proportionality and Autonomous Weapons Systems"; Sparrow, "Building a Better Warbot."
[139] See scenario 1 'Killerbot' in Appendix A for the full description of the situation.

When it comes to targeting,[140] the necessity principle in practice often encompasses two questions that must still be answered after other obligations under IHL are shown to have been complied with:

a) Is the action required for direct military advantage, or, as the US Air Force puts it, "required to quickly and efficiently defeat the enemy?"[141]

b) Is the target of the action a valid 'military objective'?[142]

In the case that question b is answered with a 'no', the principle of distinction would per definition be violated, making the action impermissible under IHL. In judging the extent to which certain military functions could be performed by RAS, especially in the case of (semi-)autonomous weapons, it is therefore crucial that the system in question possesses the ability to distinguish valid military targets (see question b). This is often a context-heavy question—and once it is answered, the even more environment-dependent issue of direct military advantage will likely remain. This will require forms of human–machine teaming at least in the near future, with humans' experience, ability to draw from varying contexts, and creativity in assessment remaining relevant in decisions on the use of force.

Even more so than proportionality, the principle of military necessity remains controversial. Military necessity within IHL recognizes (gaining significant advantages in) winning a war as a legitimate consideration towards the use of force and legitimizes collateral damage to an extent. The principle can be seen in both a more permissive and a more restrictive light. A relatively widely accepted view falls in the middle, namely that the principle of necessity is itself, in general, a permissive principle, while the later-discussed principle of humanity has a necessary limiting function that counterbalances this.[143]

The issue of military necessity was raised frequently in scenario 4 of the expert session.[144] The experts weighed whether the deployment of a certain autonomous defensive system at the border that could identify and down aircraft deemed to be a threat was necessary, due to the risk of two types of accidents that could occur. The first was the accidental downing of non-military aircraft that could violate the airspace, and the second was the downing of military aircraft that could have flown near the border but were actually only flying towards the system to test its parameters and

[140] The word 'targeting' does not refer to only the (kinetic) action against a target but rather, it indicates the larger military decision-making process. The various phases in this process are set out for example in Ekelhof, "Autonome Wapensystemen: Wat We Moeten Weten over de Toepassing van Het Humanitair Oorlogsrecht En de Menselijke Rol in Militaire Besluitvorming," 199–202; Ekelhof, "Lifting the Fog of Targeting: 'Autonomous Weapons' and Human Control through the Lens of Military Targeting."

[141] "Annex 3-60 – Targeting. Appendix A: Targeting and Legal Consideration. Basic Principles of the Law of War and Their Targeting Implications," 89.

[142] According to Article 52 of Additional Protocol I to the Geneva Convention, military objectives are "those objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage."

[143] Melzer, "Keeping the Balance between Military Necessity and Humanity – a Response to Four Critiques of the ICRC's Interpretive Guidance on the Notion of Direct Participation in Hostilities," 833; Schmitt and Thurnher, "'Out of the Loop': Autonomous Weapon Systems and the Law of Armed Conflict."

[144] See Scenario 4 'Back to basic' for the full description of the situation.

the willingness of the defending state to engage the aircraft. One set of experts concluded that using the system was the most ethical approach to defense—if the intentions and the system's parameters had clearly been communicated to the adversary's higher command. This way, the defending force had explicitly delineated its position on the engagement of targets.

The obligation to ensure military necessity before resorting to the use of force is one that, in the context of AWS debates, brings us back to the concept of human control. RAS can reduce the amount of time needed to come to conclusions or to make decisions—something that is becoming increasingly important both on the physical battlefield and in the cyber domain.[145] The way in which RAS can speed up decision-making is a double-edged sword, however. The more this speed surpasses what human reasoning is capable of, the more military technologies may shift from being largely diagnostic or descriptive to becoming more predictive or even prescriptive. There are two separate types of reasoning on which human decisions tend to be based: deliberative and automatic.[146] The high speed at which RAS could perform analyses in order to keep up with technological developments in warfare increasingly requires humans to supervise or make decisions on whether to follow a system's 'judgment' using automatic reasoning rather than longer, more weighed deliberative reasoning. The surpassing of human cognitive abilities together with the knowledge that lives may be in danger from a mistake, can result in a sense of urgency that further affects human judgment on whether or not the correct conclusions have been drawn by the system. This may effectively leave certain originally human decisions up to machines, and ultimately it may warp what is considered an absolute necessity or an 'imminent threat'. This is especially relevant to decisions concerning whether or not to use kinetic force, whether defensively or offensively.[147]

Distinction

The third principle to be adhered to is the notion that there must be a distinction between legitimate (e.g., active military and combatant) and non-legitimate targets (e.g., civilians, civilian objects, surrendering soldier, or medical staff). This distinction lies at the core of the regulation of hostilities.[148]

For the purposes of this paper's topic, the principle implies that there must be a sufficient level of certainty that RAS can ensure distinction between different 'types' of actors in a potential zone of action. In practice this means that militaries may only use weapons that can distinguish valid targets from civilian or protected targets. Presently, lacking autonomous weapons with such advanced

[145] "Reframing Autonomous Weapons Systems."

[146] Noel Sharkey does this based on human psychology research, which "divides human reasoning into two types: (i) fast *automatic* processes needed for routine and/or well tasks like riding a bicycle or playing tennis and (ii) slower *deliberative* processes needed for thoughtful reasoning such as making a diplomatic decision." See Sharkey, "Guidelines for the Human Control of Weapons Systems," 2.

[147] Schwarz, "The (Im)Possibility of Meaningful Human Control for Lethal Autonomous Weapon Systems."

[148] Netherlands Advisory Council on International Affairs, "Autonomous Weapon Systems: The Need for Meaningful Human Control."

abilities, militaries must have human decision-making in place at all points in the targeting process where it is necessary to ensure the principle of distinction is upheld.

On the topic of distinction there is much debate surrounding the abilities of RAS, as even for humans it can be challenging to distinguish between combatants participating directly in hostilities and the locations or buildings associated with them.[149] Presently, existing AWS can only 'know' the difference between military targets and civilian objects under particular circumstances and in particular environments. However, training with systems that have learning abilities may be able to enhance their adaptivity and the possibility to prepare for far more different scenarios than initially programmed. Furthermore, similar to many other technologies or weapons, RAS are to be deployed for specific contexts and aims, as not all systems or programs fit all aims. At least for now though, RAS identify targets and warning signs based on certain, pre-programmed criteria, whereas conflict situations are unpredictable, and the identification of combatants does not usually adhere to easily programmable criteria. It is therefore unlikely that in the foreseeable future it will be possible to have a (weapon) system autonomously identify valid military targets.[150]

Humanity

The last of the four principles in IHL discussed here, is the principle of humanity. While in and of itself being a principle that enhances—and often limits—the aforementioned three, there are a number of elements that make humanity a standalone principle.[151]

In an important International Court of Justice (ICJ) advisory opinion, the prohibition of unnecessary suffering is made explicit.[152] The notion that the "employment of arms which uselessly aggravate the sufferings of disabled men, or render their death inevitable" would be "contrary to the laws of humanity" has been a core tenet of international law governing hostilities dating back as far as 1868.[153] It is therefore crucial in the development of RAS to keep this overarching principle in mind.[154]

The main point of debate surrounding the principle of humanity that is relevant to this paper, is whether or not military RAS will violate the 'Martens Clause' under IHL.[155] This clause can be found

[149] For countries' description of combatants, see "Customary IHL - Practice Relating to Rule 14. Proportionality in Attack." For there to be direct participation in hostilities, there are three criteria: "a threshold of harm, a causal link between the act and the harm, and a connection to one of the parties to an armed conflict", Netherlands Advisory Council on International Affairs, "Autonomous Weapon Systems: The Need for Meaningful Human Control," 25.

[150] Netherlands Advisory Council on International Affairs, "Autonomous Weapon Systems: The Need for Meaningful Human Control," 24–25.

[151] Davison, "A Legal Perspective: Autonomous Weapon Systems under International Humanitarian Law."

[152] Legality of the Threat or Use of Nuclear Weapons (Advisory Opinion of 8 July 1996), 1996 Reports paragraph 78.

[153] Declaration Renouncing the Use, in Time of War, of Explosive Projectiles Under 400 Grammes Weight.

[154] Declaration Renouncing the Use, in Time of War, of Explosive Projectiles Under 400 Grammes Weight.

[155] See Hughes, "No, Autonomous Weapon Systems Are Not Unlawful under the Martens Clause"; Docherty, "Banning 'Killer Robots': The Legal Obligations of the Martens Clause"; Asaro, "Jus Nascendi, Robotic Weapons and the Martens Clause."

in different forms throughout various IHL treaties, but is most often quoted as follows, from Article 1(2) of Additional Protocol I to the Geneva Conventions:

> *In cases not covered by this Protocol or by other international agreements, civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and from the dictates of public conscience.*[156]

There are different interpretations of this, more often than not depending on the status of the interpreting actor and their stakes in a conflict.[157] The limiting interpretation is that the Martens Clause can be a legal argument—of customary international law—for the prohibition of certain actions or, in the case of RAS, certain weapons.[158] The opposing, permissive interpretation is that this clause is relatively insignificant, and simply functions as reaffirming signatories' acknowledgment of governance by customary international law. The third, more middle-ground reading of the clause is that it can function as a legal argument to argue the illegality of certain systems, either alone or in conjunction with other legal arguments, but is itself not inherently a prohibition of any action or weapon per se.[159]

There has been debate over the extent to which the Martens Clause applies in the case of military RAS. On the one hand, the clause may apply to AWS because these are not addressed specifically by international law.[160] On the other hand, although RAS nor AWS are specifically mentioned, their use and the limits to this use are dictated by IHL and the weapons review regulations of customary international law.[161] These different viewpoints will be expanded upon in 4.2, as they are part of key discussions on AWS developments' effects on human dignity. Most importantly though, the Martens Clause prevents "the assumption that anything not explicitly prohibited is permitted," and thereby has the ability to act as legal grounds for various policy directions taken in "new situations and new means and methods of warfare."[162] The notion from the Martens Clause that the development of new technologies like RAS depends in part on the dictates of public conscience, makes societal debate an important consideration for the militaries of democratic nations in deciding whether or not, and how, to develop military RAS.

4.2 Dignity in the use of AWS

Linked to the principles of IHL discussed above, as well as to the earlier topic of meaningful human control, a debate exists concerning the basic understanding of what is most 'ethical' or 'humane'.

[156] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I).
[157] Sparrow, "Ethics as a Source of Law."
[158] Docherty, "Losing Humanity."
[159] Cassese, "The Martens Clause: Half a Loaf or Simply Pie in the Sky?"
[160] Docherty, "Losing Humanity"; Docherty, "Heed the Call."
[161] Press, "Of Robots and Rules: Autonomous Weapon Systems in the Law of Armed Conflict."
[162] Davison, "A Legal Perspective : Autonomous Weapon Systems under International Humanitarian Law," 8.

While future capabilities of RAS might make certain missions easier for the military, or may end up helping save lives in conflict,[163] their use can still be considered as diminishing certain inherently human aspects of conflict. There is, however, no set definition of what the most dignified way is to go about preventing suffering, making it difficult to identify a widely accepted explanation of what is or is not 'humane' and 'dignified'. The two sides arguing this debate tend to operate on a different level of understanding of the notion of 'humane': on the one hand, it is said that decisions of life and death are inherently human and delegation thereof to a machine would be per definition inhumane; on the other hand, it is said if human suffering can be reduced, then not doing so would be inhumane. To illustrate this with a hypothetical: if using a certain type of RAS were to decrease the number of civilian deaths from twenty to ten in a particular conflict situation, one could either argue that the inherent fact that RAS use led to ten people dying is counter to human dignity, or one could argue that RAS made the situation more dignified because IHL was better observed, as ten less people died than otherwise would have been the case.

The following two sub-sections present the arguments for both sides of the debate on human dignity in warfare. The first sub-section presents the argument that the introduction of RAS will undermine dignity, while the second sub-section discusses how RAS may strengthen the position of human dignity in warfare. This discussion is presented as a debate in order to lay out the key arguments that have been presented to support or oppose development and deployment of military RAS in both the near and the more distant future.

Perspective One: Undermining Dignity

As the realm of machine learning expands, militaries stand to gain, both defensively and offensively, in terms of speed and efficiency. Automated defenses are not particularly new—think back to the earlier example of Israel's Harpy system—but future RAS may go further than protection and venture into the realm of counter-attacks.[164] Not only this, but machine learning algorithms will increasingly be employed to help inform decisions on resorting to the use of force internationally.[165] This of course raises issues about whether such systems provide a reliable analysis, and to what extent a human operator can question or overrule the recommendations produced through AI systems. As the value of such algorithms resides in their speed and the possibility to react immediately if it is deemed necessary for self-preservation, "the temptation to rely on the algorithm alone to guide decision-making [...] will be powerful."[166]

As will be discussed in the following paragraphs, primary arguments against RAS in the context of human dignity pertain to decisions on the use of force, distance, explainability, and threats to peacebuilding.

[163] Arkin, "Lethal Autonomous Systems and the Plight of the Non-Combatant."
[164] Deeks, Lubell, and Murray, "Machine Learning, Artificial Intelligence, and the Use of Force by States."
[165] Deeks, Lubell, and Murray.
[166] Deeks, Lubell, and Murray, 10.

An issue brought up most often is that the barriers to using force may be reduced by the increased use of RAS, forming a problem for the general rule that the use of force should be a last resort, intended as self-defense.[167] This lowering of thresholds to violence could happen in different ways. For one, if there are fewer human lives endangered on the side of the attacker, the risk posed by the possible outcomes of an attack is lowered. Alternatively, it could happen that an early warning system is spoofed[168] by another (state) actor into 'recognizing' imminent danger when certain pre-programmed boxes are ticked, thereby setting off a counter-response.[169]

On the issue of distance, recent decades' development in military involvement abroad and the changing nature of this type of conflict form an interesting backdrop to the further integration of RAS in the military. In the case of the US, the international operations it has carried out over the past years, with less 'boots on the ground' and a larger role for military contractors, enabled the country to "maintain the appearance of a small military footprint with minimal risk of harm to US troops."[170] Throughout interventions, from Libya to the latest intervention against ISIS in Iraq, this has paved the way for the legal groundwork to claim that having less troop casualties makes military involvement a more valid option in the US' foreign policy.[171] As more functions of soldiers can be performed at a distance and/or by an autonomous system, the argument on a national level that military action abroad is more acceptable if the country's own troops are in less danger, is strengthened. The fear here is that this physical and moral distance from the battlefield or conflict zone will further lower the barrier to the international use of force. The lines drawn to regulate to the use of force are further blurred by the prevalence of involvement by proxy; asymmetric forces; the lack of clarity on which actors or parties are involved; whether hostilities constitute a full-blown war; and where the geographical 'border' of this war should be. If a situation cannot be classified as war, then per definition IHL cannot be upheld: IHL governs warfare.

Aside from the above, the distance between the operator and the target is another point of concern—not dissimilar to discussions seen during the deployment of drones or even the initial use of airpower.[172] There is an assertion that human dignity is undermined if machines effectively have the last say in who lives or dies—be it on purpose or by accident. With the increased use of automation and autonomization, two forms of distance have a possible impact on operations: institutional and physical.[173] As for the former, operating within an ethical armed force is not just about outcomes, but also about the process that led there. Fully autonomous weapons "would lack

[167] UN Charter Article 51 ; ICJ, Nicaragua Case (Merits), para 191: only the most grave forms of attack qualify; para 176: "self-defence would warrant only measures which are proportional to the armed attack and necessary to respond to it."

[168] It appears that not only other actors hacking a country's system is a realistic risk, but so too is 'spoofing', or the tricking of system algorithms, often by mimicking patterns known to set off a certain reaction in the system. Extensive experiments with AI image-classification algorithms have shown that these systems are easily tricked with relatively small deviations from standardized representations. For example, one such experiment saw a turtle continuously being misidentified as a rifle. See Klare, "Autonomous Weapons Systems and the Laws of War."

[169] Deeks, Lubell, and Murray, "Machine Learning, Artificial Intelligence, and the Use of Force by States."

[170] Dickinson, "Drones, Automated Weapons, and Private Military Contractors," 111.

[171] Dickinson, "Drones, Automated Weapons, and Private Military Contractors."

[172] Banta, "'The Sort of War They Deserve'?"

[173] Feickert et al., "U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI): Considerations for Congress."

the human judgment necessary" to ensure all agreements and obligations are followed through in a way that can be justified after the fact.[174] And as for the latter, even with a human closely monitoring or making certain decisions, the physical distance created by (semi-)autonomous systems can also lead to a "moral distance as the face of the opponent becomes less visible, which eliminates the moral–psychological barrier for killing."[175]

The explainability of the decision-making process—as described in Chapter 3.1—is closely linked to the previous point on the distance of the human from the battlefield. Using RAS not just on an operational or tactical level, but incorporating such complex systems into wider decision-making processes will add to the already prevalent transparency issues that the military has toward the public—and transparency is a "key element in enabling society to have the right amount of trust and confidence in the operations of an AI system."[176] With a lack of transparency into decision-making comes less scrutiny of the quality and interpretability of machine-produced recommendations or predictions decisions were based on, and therefore it may also lead to a lessened sense of responsibility.[177] Aside from the possibility that for this reason a state may be less willing to explain machine reasoning behind decisions, a very real possibility is also that a state may unable to explain it—a possibility with far more consequences for the position that human dignity considerations hold in policy-making.[178]

Lastly, we come to arguments pertaining to the post-war effects of military RAS on human dignity. While increasingly the world sees conflicts of various intensity with no clear start or end, the feasibility of peacemaking and peacebuilding should nevertheless always be on the mind of involved states. Visible use of RAS may undermine counterinsurgency efforts intended to stabilize a region, in a way similar to how civilians can react to or fear the use of force by armed drones. A military's intention may be good, but all that is heard or seen on the ground is the noise of an overhead drone and the resulting destruction.[179] At this point in time, for similar reasons, RAS "won't help win the hearts and minds of the occupied or vanquished," and may make it more difficult to achieve lasting peace.[180] RAS use can be perceived on the ground as a lack of commitment if used by a state making peacebuilding efforts. Aside from this, it may also hamper partnership efforts. This is a point especially relevant to forces—like the RNLA[181]—which have an approach to peacebuilding that takes into account the importance of understanding the area of operations and the local sensitivities that can make or break the success of such missions. RAS are incapable of developing

[174] Davison, "Autonomous Weapon Systems"; "Killer Robots - Learn."
[175] Verdiesen, "Agency Perception and Moral Values Related to Autonomous Weapons," 14.
[176] Charisi et al., "Towards Moral Autonomous Systems."
[177] Johansson, "Ethical Aspects of Military Maritime and Aerial Autonomous Systems."
[178] Feickert et al., "U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI): Considerations for Congress."
[179] Khan & Gopal, "The Uncounted".
[180] Lin, Bekey, and Abney, "Autonomous Military Robotics: Risk, Ethics, and Design."
[181] van der Lijn and Ros, "Peacekeeping Contributor Profile: The Netherlands."

or replacing the personal relationships with the local population that are generally necessary for trust and, by extension, for a successful mission.[182]

Perspective Two: Aiding Dignity

Several points run counter to the above discussed argument that military RAS would lower barriers to the use of force.[183] The first is the fact that as the speed of (cyber) attacks steadily increases, so does the speed with which a response must be readied. Machines could calculate endless different options, outcomes, and consequences at far greater speed and accuracy than humans, and could in the future therefore improve both decisions on self-defense and counter-attacks, as well as on legitimate targets or intensity of attacks. This is of course still dependent on the possibility of satisfying the requirements of proportionality, necessity, and distinction. Another argument is that the public knowledge that certain states have certain capabilities may actually work as a deterrent.[184] This was also noted in the scenario 4 of the expert sessions, where the autonomous system and its parameters that were communicated to the adversary's higher command limited the strategic choices of the adversary, as a defensive AWS is more likely to be consistent in its behavior than a human operator.

Another common argument is that distance will also lower the threshold to using force. However, through the use of a semi-autonomous system operated at a distance, reduced stress levels and increased evaluation time may allow a human operator the opportunity to make better informed and as a result, more ethical decisions.[185] It can be argued that if a reliable and IHL-observant autonomous system is developed, for any relevant type of operation, it should perhaps "not only be regarded as morally acceptable but also [...] ethically preferable over human fighters" if unethical situations can thereby be avoided.[186]

An additional reason that RAS could in the future add to human dignity, is the lack of human shortcomings, such as fatigue or emotions clouding judgment. It is possible to embed a mission's rules of engagement (ROEs) in a system, and once technological development is far enough that RAS can observe IHL, this could prevent mistakes or even war crimes that may otherwise arise under the strenuous circumstances soldiers are often in.

With respect to the discussion on the difficulty of explaining the actions of RAS, results from machine-produced recommendations may make it easier for states to lay out the rationale that led to a certain outcome—rather than a decision being a commander's intuition.[187] A difficulty here

[182] Marchant et al., "International Governance of Autonomous Military Robots," 2889.
[183] Feickert et al., "U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI): Considerations for Congress."
[184] Deeks, Lubell, and Murray, "Machine Learning, Artificial Intelligence, and the Use of Force by States."
[185] Davison, "Autonomous Weapon Systems"; Arkin, *Governing Lethal Behavior in Autonomous Robots*, 47–48; Strawser, "Moral Predators"; Leveringhaus, "Autonomous Weapons Mini-Series."
[186] Etzioni and Etzioni, "Pros and Cons of Autonomous Weapons Systems," 74.
[187] Deeks, Lubell, and Murray, "Machine Learning, Artificial Intelligence, and the Use of Force by States."

is that much of what goes on in algorithmic functions is a 'black box', and while some systems log decision-making and errors, with increasingly complex systems it can be difficult to establish precisely how some results were reached. This topic was discussed in more depth in Chapter 3.1.

Overall, there is still a long way to go before machines could autonomously satisfy IHL requirements. However, certain examples can show that there are rapid developments in this field, sped up by the fusion of R&D in civil and military spheres. Such overlap may prove necessary to keep up with the way in which war is changing. The world is becoming increasingly urbanized, with over two thirds of the world projected to live in cities by 2050,[188] creating complex security problems for local governance that can escalate and become local or regional conflicts.[189] At the same time, the trend of conflicts taking place in cities is only set to continue,[190] and the packed, hard to navigate, and 'easy-to-hide-in' nature of cities, means that the principles of IHL will only become harder to adhere to for soldiers in pressed positions.[191] This means that states developing RAS have a great interest in further research into RAS capabilities that will improve their militaries' ability to withstand the IHL test in the most complicated of environments.[192] The US, for example, uses war games in fictitious cities to run through scenarios and establish what types of technology it will need to get through the coming decades, knowing that all war-fighting functions "are complicated and challenged by the compartmentalized terrain that's present in the urban environment."[193] RAS may provide a crucial tool enabling militaries to have the intelligence and preparation needed to better assure all precautions are taken and all IHL principles are upheld in the complicated modern war scenarios.

4.3 Summary

Crucial to upholding human dignity in conflict is adherence to and respect for International Humanitarian Law. The most important principles that RAS should be able to respect are proportionality, military necessity, distinction, and humanity (often known as the prevention of unnecessary suffering). Developments of existing AWS are making strides in this regard. They can 'distinguish' to a certain extent the difference between military targets and civilian objects, albeit only under particular circumstances. Training with systems that have learning abilities may be able to enhance their adaptivity and the possibility to prepare for far more different scenarios than initially programmed. RAS can greatly reduce the amount of time needed to come to conclusions or to make decisions. At the same time, however, surpassing human cognitive abilities strongly

[188] United Nations, Department of Economic and Social Affairs, Population Division, "World Urbanization Prospects: The 2018 Revision."

[189] Horowitz, "Joint Blog Series: Precautionary Measures in Urban Warfare: A Commander's Obligation to Obtain Information."

[190] "Preparing for More Urban Warfare."

[191] Horowitz, "Joint Blog Series: Precautionary Measures in Urban Warfare: A Commander's Obligation to Obtain Information."

[192] A way RAS can improve militaries' functioning in such difficult environments is, for example, a tool like Hivemapper. This "creates 3D maps from videos captured by drones, aircraft, and ground vehicles" as a way of "using machine learning tech to augment human analysts." See Weisgerber, "What's in the House NDAA?; Pentagon's 3D-Mapping Service; New Marine One, Weed Whacker; and More."

[193] Calloway, "Army Wargames Shape the Future of Urban Warfare"; Musgrave, "Inside 'Liberty City,' Homeland Security's Site for Testing Urban Drones."

affects the mitigating role of human judgment in assessing whether or not the correct conclusions have been drawn by the system. Even so, RAS capabilities can also improve militaries' ability to navigate the complicated environments in which modern conflicts tend to be fought, by improving intelligence, logistics, evacuation and other capabilities crucial to effective mission functioning. At the same time, for the foreseeable future, it will not be possible to have an offensive (weapon) system that can autonomously distinguish to the standards of IHL and weigh all relevant context as well as a human.

The debate on the effects of integrating RAS into the armed forces is one with a wide range of arguments. The ongoing academic debate is important to inform policy-makers, but in the end, decisions are made politically. Meanwhile, Machine Learning and Artificial Intelligence are becoming more common in shaping decisions in military contexts. It is also likely that intelligent system design will continue to become more intricately woven into the fabric of decision-making—not only on the battlefield, but also in the policy-making world that governs it. In the future this development will, however, continue to make the attribution of wrongful use of force. In general, RAS may provide a crucial tool enabling militaries to improve their intelligence and preparations can help assure that the necessary precautions are taken and IHL principles are upheld in the complicated modern war scenarios.

5. Responsibility and Accountability

The final component of ethics in the context of this study concerns the establishment of accountability and responsibility before, during, and after the deployment of RAS in a military context.

This chapter will re-examine the three perspectives of meaningful human control (MHC) introduced in Chapter 3, namely the life cycle, sub-system and OODA loop perspectives (See Figure 3). This approach highlights all the elements to MHC, and is useful in illustrating the numerous actors involved with RAS and all the points at which they may be (partially) responsible for certain courses of action. The approach also exemplifies why it is so important in the context of ethics to think about accountability beforehand, in order to ensure it is clear who shares responsibilities throughout the long and relatively fragmented chains of R&D and usage that typify military technology.

The discussion on military RAS often seeks to compare humans and machines. We can take as an example "the difference between a pilot flying an airplane on autopilot and an airplane with no human in the cockpit at all."[194] The relevant question for this chapter on accountability and responsibility does not dwell on whether the former or the latter way of flying is 'better'. The question is rather, in the case of system or human failure, what is the amount of damage caused by their failure that will be deemed acceptable by the military or society? How can this be assessed early on and how can the risks associated with flying on autopilot or autonomously be mitigated, in line with that baseline of ethical standards?

There are currently some obstacles to determining responsibility and establishing accountability for activities involving RAS.[195] First, sub-chapter 5.1 goes into the current ways in which accountability may apply in the case of RAS from a legal point of view,[196] especially in cases where wrong was done by mistake or otherwise. The fact that the intent behind actions is a part of the legality thereof presents a difficulty, as it is not always clear whose intention should be reckoned with when it comes to the deployment of RAS, nor how this intent could be sufficiently established in a legal sense in the first place. Secondly, 5.2 discusses whether existing legal and accountability frameworks are sufficient to safeguard society's ethical standards, and it goes into the extent to which there may be a legal accountability gap at this stage and in the future. Following on from

[194] Scharre, *Army of None: Autonomous Weapons and the Future of War*, 193.

[195] Schmitt and Thurnher, "'Out of the Loop': Autonomous Weapon Systems and the Law of Armed Conflict."

[196] In legal terms, 'responsibility' refers to a duty to act with due diligence. 'Accountability' refers to "the process aimed at a [...] public assessment of [...] conduct in a given case in order to evaluate whether this conduct was required and/or justified" based on established responsibility. Finally, the term 'liability' follows this, and refers to the attachment of legal consequences to said conduct.

this, sub-chapter 5.3 looks at where there is room for developments to better address the challenges of accountability for RAS, and addresses how to understand responsibility and accountability in a non-legal, institutional way.

5.1 Current Frameworks of Legal Accountability for Military Practices

Integrating RAS into military operations may erode moral responsibility, as repercussions for IHL non-compliance requires legal individual and/or group accountability. If one would seek some form of accountability for certain outcomes of RAS' actions or decisions, there must be shown to be a link between the outcome of a RAS-dictated action and the intent of those responsible for its development and/or operation.[197] First, the discussion on legal personhood of RAS will be set out briefly, given recent years' developments in this field, especially for European nation-states.[198] After this, the two primary means through which one could establish legal accountability and liability are discussed, namely state and criminal responsibility, as well as the shortcomings tied to these legal frameworks in the governance of RAS. Lastly, a number of alternative bodies of law have been floated as useful to incorporate or take from in establishing a legal framework to deal with possible RAS cases in the future.

Legal Personhood and Autonomous Systems

With RAS at its present stage of development, in the ethical and legal sense, responsibility and accountability for their actions fall upon humans.[199] While legal personhood already exists for international organizations and companies, for example, the European Parliament has suggested to the European Commission to consider extending this to a form of legal, "electronic" personhood for robots.[200] This suggestion has received mixed reviews.[201]

There are several points to do with this notion of legal personhood for autonomous synthetic entities. First and foremost,

> *[t]he basic provisions for a legal person are: 1. that it is able to know and execute its rights as a legal agent, and 2. that it is subject to legal sanctions ordinarily applied to humans.*[202]

[197] International Committee of the Red Cross ICRC, "Ethics and Autonomous Weapon Systems"; Feickert et al., "U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI): Considerations for Congress."Switzerland","event-place":"Geneva, Switzerland","abstract":"In the view of the International Committee of the Red Cross (ICRC

[198] See Committee on Legal Affairs, "Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))."

[199] Shilo, "Speaking of Responsibility: Autonomous Weapon Systems, State and Individual Responsibility"; Lin, Bekey, and Abney, "Autonomous Military Robotics: Risk, Ethics, and Design."

[200] The suggestions made by the Parliament include a definition of what characteristics would constitute 'smart' autonomous robots. See Committee on Legal Affairs, "Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))."

[201] See e.g., Bryson, Diamantis, and Grant, "Of, For, and By the People: The Legal Lacuna of Synthetic Persons"; Vincent, "Giving Robots 'Personhood' Is Actually about Making Corporations Accountable."

[202] Bryson, Diamantis, and Grant, "Of, For, and By the People: The Legal Lacuna of Synthetic Persons."

Second, 'legal personhood' is a technical term that does not necessarily imply somehow viewing robots as inherently human or ethical actors. Rather, the term gives way to any number of rights and obligations, because it means that a legal system addresses its rules to the actor or entity.[203] While much discussion on the European Parliament's report focused on this notion of electronic personhood, the aim of the report's recommendations was more to ensure that establishing "a causal link between the harmful behavior of the robot and the damage suffered by the injured party" could become sufficient to be able to claim compensation from a company. The extension of the term 'legal personhood' is largely a legal tool of convenience within civil law, in the same way legal personhood is given to a company in order to provide them with legal rights as well as obligations.

The autonomous robots envisioned in this Parliamentary text were for civil use rather than military. However, anticipating future risks posed by increased autonomy and addressing the liability gap that may arise if no legislative care is taken to address this topic, is equally relevant for military RAS. More on civil law instruments that could be relevant for military RAS will be covered further on in this chapter.

State Responsibility

Responsibility for state wrongdoing is established on the basis of Article 2 of the International Law Commission's (ILC) Draft Articles on the Responsibility of States for Internationally Wrongful Acts, which dictates that

> *there is an internationally wrongful act of a State when conduct consisting of an action or omission: (a) is attributable to the State under international law; and (b) constitutes a breach of an international obligation of the State.*[204]

However, the state itself as an entity is made up of the people that represent it, meaning "an 'act of the State' must involve some action or omission by a human being or group [...]."[205] Furthermore, the "only conduct attributed to the State at the international level is that of its organs of government, or of others who have acted under the direction, instigation or control of those organs, i.e., as agents of the State."[206] At this point a distinction can be made between on the one hand agents of the state that take orders within an explicitly established command structure, and on the other hand agents of the state that make an individual decision that results in wrongful act, outside of an 'effective command and control' structure.[207] The latter will be discussed further in the criminal responsibility section. In conclusion, a breach of international law has a human link and requires the presence of humans, which at the state level manifests itself through agents of the state.

[203] Bryson, Diamantis, and Grant.
[204] International Law Commission, "Responsibility of States for Internationally Wrongful Acts."
[205] International Law Commission, 35, paragraph 5.
[206] International Law Commission, 38, paragraph 2.
[207] "Killer Robots and the Concept of Meaningful Human Control."

States have the duty to respect and ensure compliance with IHL under Common Article 1 of the Geneva Conventions.[208] Moreover, state obligations include the regulation of companies to ensure that emerging technologies are not in violation of IHL.[209] This obligation is extended under Article 36 of the Protocol Additional I to the Geneva Convention, whereby

> *[i]n the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.*[210]

The state is therefore made responsible for testing and certifying RAS provided by domestic private contractors and/or foreign suppliers. Compliance with Article 36 requires integration of international obligations throughout most of the life cycle of RAS, from design and procurement to its adoption by the military. Developing and institutionalizing a system such as the example presented by Australia in 5.1 could be a way of ensuring that there is always state responsibility where it is required under a nation's international obligations. Without a clearer overview of where and to what extent it can be reasonably expected for a state's responsibility to lie in RAS research and development, it is difficult to hold states accountable for the consequences of their use.

Criminal Responsibility

A cornerstone of international criminal law (ICL) is the attribution individual criminal responsibility. For this body of law to be applicable, there must be criminal intent (*mens rea*) involved, or, as is generally the case for war crimes, the wrongful act must have been committed "willfully."[211] This criminal intent is what separates civil and criminal law. Someone who drives a car into a pedestrian with the intention to hurt or kill will be liable under criminal law. However, someone who loses control of their car, cannot brake in time, and hurts a pedestrian in the ensuing accident, will be liable under civil law and will most likely end up paying monetary damages as a result.

ICL cannot be applied to RAS directly. As RAS have no consciousness, no criminal intent can be established. One could also question whether this would even make sense in the first place, as the purpose of ICL is to establish willful wrongdoing and appropriate punishment, and there is not much effectiveness in applying human punishments to machines. Hence, for there to be criminal accountability under certain circumstances, an individual, or group of individuals, will need to be

[208] "Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field. Geneva, 12 August 1949."

[209] This type of obligation is known as a 'positive obligation', meaning states must make an active effort to ensure compliance with the law, such as adopting (new) measures to uphold it. A 'negative obligation' means simply to refrain from certain acts that would violate the law in question. These terms are most used in International Human Rights Law.

[210] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I).

[211] Crootof, "War Torts: Accountability for Autonomous Weapons." "Willfully" here means that someone must have acted either intentionally or recklessly, see Sandoz, Swinarski, and Zimmermann, "Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949."

held responsible. Assigning individual responsibility under ICL for the use of RAS/AWS will be difficult, however. This is largely due to the still inadequately agreed upon concept of meaningful human control, as well as the increase in decision-making by or based on machines coupled with the degree of unpredictability that still exists at the current stage of technology's development.

Criminal responsibility can—in theory—be attributed to individuals in all phases of the RAS life cycle, from programmers to operators. In practice, it would prove complicated to pinpoint one or a few culprits among the many people involved throughout RAS life cycles. A RAS criminal case would be relatively 'easy' to solve if it were possible to trace back the machine-produced wrongful act to certain people, for example if there were deliberately incorrect handling by a programmer, or recklessness on the part of an operator. However, there are two key difficulties here. First, there is the sheer number of people tasked with building an algorithm, building RAS hardware, developing training and evaluation, and all other steps involved in RAS research and development. Second, most issues in applying ICL to RAS operations arise from oversight or mistakes, rather than intentionally wrongful acts, meaning no one can be held directly liable. This fact, that RAS may result in serious violations of IHL without human intention as the main driving or facilitating factor, seriously hampers attaining justice after wrongful acts. The diffusion of responsibility means it becomes more likely that no one will be punished even in light of a mistake with lethal consequences as a result of the use of RAS.

There are various doctrines within ICL, which cannot all be discussed here, but some of which may be relevant in assigning responsibility in a military context. Not all ICL doctrines require every participant in a crime to have intended this crime, but they are still all "premised on the notion that there is at least one individual who did possess the requisite intent."[212] One of these doctrines often cited as being most relevant in when it comes to military RAS is that of command responsibility. In this doctrine, a commander can be held legally responsible for the actions of a subordinate if they had "effective command and control, or effective authority and control over the forces that committed the crime."[213] In a ruling by the International Criminal Tribunal for Rwanda, it was established that the "material ability to control the actions of subordinates is the touchstone of individual [command] responsibility."[214] In other words, this responsibility is conditional upon not only the subordinate(s) intent to commit a wrongful act, but also upon the commander's material ability—or lack thereof—to actually prevent and punish the commission of the offense.[215] If it is proven that a commander was not realistically in a position to prevent or punish the actions of RAS/AWS, this means it is unlikely that the commander would be criminally liable. This is the case since criminal responsibility is attributed after the fact (*ex post*), while the use of AWS is permitted

[212] Dickinson, "Drones, Automated Weapons, and Private Military Contractors," 116.
[213] "Killer Robots and the Concept of Meaningful Human Control"; Galand, Hunter, and Utmelidze, "International Criminal Law Guidelines: Command Responsibility," 65.
[214] Prosecutor v Kavishema paragraph 229.
[215] Shilo, "Speaking of Responsibility: Autonomous Weapon Systems, State and Individual Responsibility"; Mucic et al ("Celebici") paragraph 378.

before the event (*ex ante*) under the CCW and in compliance with IHL.[216] Moreover, in the case of supervised autonomous systems with the possibility of human override, rather than see their well-intentioned intervention resulting in a negative outcome, an operator may instead prefer to benefit from plausible deniability after inaction. Thus, there is an adverse incentive for the operator overseeing the RAS not to intervene if mistakes may be met with criminal liability afterward.[217]

The issue with attempting to assign criminal responsibility, within any of the doctrines considered, is that it is inherently tied to the individual and their intentions. There are increasingly lengthy research and development processes for RAS, as well as more diffusion of tasks in these processes among government, military, and private sector actors, meaning that there are almost never specific individuals solely responsible for the consequences of RAS deployment. Moreover, where exactly state responsibility starts and ends is presently unclear for parts of the RAS life cycle, and individual criminal responsibility is often not applicable. Altogether, the existing legal frameworks surrounding responsibility for, and dealing with, mistakes as well as wrongful actions as a result of the use of military RAS appear to have limitations.

Alternative Bodies of Law

Several bodies of law have been brought forward in attempts establish responsibility as well as make remedies for wrongful acts by AWS possible. The latter is relevant for cases where no human criminal intent can be established, but there was still wrongdoing. One such body of law is contract law, which has also been considered for regulating private military and security contractors (PMSCs) more generally.[218] The Montreux Document, for example, suggests that States "include contractual clauses and performance requirements that ensure respect for relevant national law, international humanitarian law and human rights law by the contracted PMSCs."[219] Something could be said for contract law's possibility to circumvent jurisdictional obstacles to regulating wrongful acts in militaries dependent on private contractors, including for their RAS. Contract law may be a way to force private contractors to adhere to the norms of public international law.[220] There are issues with using private law to remedy public injustices, however. For one, regulating war crimes through the lens of contract law creates denial: "harm is cognitively reframed and then allocated to a different, less pejorative class of event"[221] until "a human rights violation is the same as a breach of contract."[222] Another issue is that all responsibilities or obligations are limited to precisely what is included in a contract's terms. Diving into the minutiae of one's precise obligations and the exact

[216] Shilo, "Speaking of Responsibility: Autonomous Weapon Systems, State and Individual Responsibility"; Bo and Woodcock, "Blog: Lethal Autonomous Weapons, War Crimes, and the Convention on Conventional Weapons."
[217] Chehtman, "New Technologies Symposium."
[218] Liu, "Contract Law as Cover."
[219] "The Montreux Document - On Pertinent International Legal Obligations and Good Practices for States Related to Operations of Private Military and Security Companies during Armed Conflict", para A.IV.14 and 15 of Part II.
[220] Dickinson, "Contract as a Tool for Regulating Private Military Companies"; Dickinson, *Outsourcing War and Peace*, 69–101.
[221] Cohen, *States of Denial: Knowing about Atrocities and Suffering*, 106.
[222] Liu, "Contract Law as Cover," 24.

terms of a contract is quite normal in private law, but this inward-looking nature conflicts with the necessity in IHL to look beyond what is on paper.[223]

A second option that has been floated introduces tort law.[224] In common law, tort law is relevant for cases where there has been wrongdoing, but no criminal intent. Whereas criminal law is set on prohibiting certain behavior, it has been suggested that tort law could offer "a means of regulating valuable but inherently dangerous activities and compensating injurious wrongs."[225] Where ICL is meant to hold individuals accountable for war crimes, "war torts" may form an added regime that could hold states accountable, circumventing the need for criminal intent. States are, in the end, responsible for making the choices that lead to the integration of RAS into the military. The idea to primarily implement a tort regime with states in mind as the responsible actors therefore has come to the foreground the most.[226] Adding on to this notion of state responsibility in tort cases is the fact that "as long as a certain type of weapon is considered lawful and its production is ordered by a legitimate entity, corporate responsibility does not pose any contentious issues."[227] This is because manufacturers are absolved of liability if the system provided meets the legal conditions of the acquiring agency at the time of order.[228] A major issue with tort law is scalability.[229] While tort is a standard procedure in domestic legal systems, there is no international regime for it. Getting a regime off the ground that would allow people or groups to essentially sue states for damages incurred due to unpredictable RAS is difficult to envision.

5.2 A Legal Accountability Gap?

One of the key challenges in the use of RAS, and AWS in particular, is that the absence of full applicability of existing legal accountability frameworks, alongside inadequate agreement over what constitutes human control over a system, creates an 'accountability gap'. By this accountability gap is meant that in cases of violation of IHL, whether accidental or intended, there may be no human or entity directly responsible. Lacking clear establishment of responsibility, there may be no accountability for the actions of the RAS.[230]

To a certain extent, human responsibility for decisions on the use of weapon systems must be retained. Accountability and liability cannot be transferred to machines themselves, and this fact should push for consideration of who holds responsibility at many different points throughout the life cycle of RAS. The fundamental problem is the existing gap in international law is based

[223] Liu, 3.
[224] A 'tort' is a civil wrong causing loss or harm, which results in legal liability for the person who committed the tortious act. Tortious acts can range from inflicting emotional distress or financial losses, to inflicting injury or invading privacy
[225] Crootof, "War Torts: Accountability for Autonomous Weapons," 1353.
[226] Malik, "Autonomous Weapon Systems: The Possibility and Probability of Accountability"; Crootof, "War Torts: Accountability for Autonomous Weapons."
[227] Malik, "Autonomous Weapon Systems: The Possibility and Probability of Accountability," 628–29.
[228] Boyle v United Techs. Corp. 487 U.S. 500, 510 (1988); Koohi v. United States, 976 F.2d 1328, 1336–37 (9th Cir. 1992)..
[229] Asaro, "The Liability Problem for Autonomous Artificial Agents."
[230] Horowitz and Scharre, "Meaningful Human Control in Weapon Systems," 8.

on the permissible use of AWS under the CCW and the criminal responsibility attributed after the unlawful act involving AWS has taken place.[231] The formed discrepancy enables the operator to cite technical issues as the cause of the incident, leaving no one accountable for the actions of the AWS. Both the operators and to some degree the programmers are further distanced from responsibility through ambiguities resulting from third-party involvement in RAS development, software update and self-learning abilities.[232] The issue was highlighted in the latest meeting at the CCW, where it was noted that "in the case of an incident involving LAWS, it was uncertain as to who would be held accountable within the chain of command or responsibility, such as the commander, programmer, or operator."[233]

What makes assigning responsibility for the actions of machines all the more challenging is the combination of RAS deployment and increased privatization of military materiel. Together, these developments have "fragmented decision-making over the use of force, rendering accountability for violations of IHL principles much more difficult to achieve."[234] Since the Nuremberg trials, IHL has been a part of the trend toward individual responsibility, and it is a crucial aspect of IHL that perpetrators are held personally responsible if they commit wrongful acts. Yet autonomous weaponry and private contractors tend not to be situated in a military command structure, bringing decision-making and the consequences for its results "outside the ordinary bureaucratic chain of command."[235] The main problem posed by this, is that it becomes far more difficult to prove that a commander has the *de facto* level of control needed to demonstrate command responsibility. While the doctrine of command responsibility is possibly a better way to assign responsibility than attempting to find individual criminal intent behind a contracted programmer, it still has no concrete way of solving cases where wrongdoing has occurred as a result of multiple actors' actions, without these actors having intended this wrong.[236] At the same time, the legal question that arose after WWII persists. This question revolves around the lack of clarity as to how far individuals truly can know what bigger picture their work is contributing to.[237] Therefore, the fact that firstly, "something more than ordinary negligence"[238] is the cornerstone of criminal responsibility; secondly, there remain a number of challenges to the establishing of state responsibility; and thirdly, other bodies of law have not yet been looked at enough in the context of RAS, means together that in light of increased privatization this body of law may prove to have significant holes, especially when it comes to RAS.

[231] Bo and Woodcock, "Blog: Lethal Autonomous Weapons, War Crimes, and the Convention on Conventional Weapons."
[232] Matthias, "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata," 181–82.
[233] Chairperson of the Informal Meeting of Experts, "Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)," para. 52.
[234] Dickinson, "Drones, Automated Weapons, and Private Military Contractors," 96.
[235] Dickinson, 115.
[236] Dickinson, 118.
[237] D Luban, legal modernism: law, meaning, and violence (Ann Arbor: U of Michigan Press, 1994) p372
[238] A Danner, J Martinez, "guilty associations" (2005), 93(1) California Law Review 77-170. P79

5.3 Addressing the Gap

As has become clear, a central question concerns where responsibility lies or should lie in the operation of RAS, in general, but most importantly in cases of unintended and/or unlawful harm.[239] Addressing this question must include the consideration of the wide spectrum of actors involved in the development and deployment of military RAS, rather than just the operators in or on the OODA loop.[240] The increasingly dominant role of the private sector in the development of RAS has propelled this conversation forward. This highlights the difficulty of both establishing responsibility and accountability throughout the full R&D and deployment chain of RAS, as well as bridging the distance from the design of a single element within a system to the final (autonomous) 'decision'.[241]

The traditional, legal way of looking at responsibility, accountability, and liability is the following. In legal terms, 'responsibility' refers to a duty to act with due diligence. 'Accountability' refers to "the process aimed at a [...] public assessment of [...] conduct in a given case in order to evaluate whether this conduct was required and/or justified" based on established responsibility.[242] Finally, the term 'liability' follows this, and refers to the attachment of legal consequences to said conduct.

There is, however, another way of looking at these three terms that could be considered within the institutional set-up of RAS integration in the military. More so outside the realm of lawyers, it has become relatively standard in the public sphere to speak of 'liability' as being explicitly rules-based and 'responsibility' as governance-based. 'Accountability' should be something that comes before both as well as after, and it refers people's ability to explain and justify their behavior at all times.[243] This is a way of defining accountability that has also become quite standard practice in business spheres, where it is described as "[t]he obligation of an individual or organization to account for its activities, accept responsibility for them, and to disclose the results in a transparent manner."[244] It can be very difficult to establish direct responsibility due to the hierarchies of public governance, making it all the more important that accountability is emphasized in fields where third-party private companies and actors have such a crucial role, as is the case for military technologies like RAS.

An example of how to institutionalize accountability and responsibility in this way was presented at the Governmental Group of Experts (GGE) meeting on the CCW in March 2019. Australia's representatives put forward a document describing how the country embeds 'control' into its weapon

[239] Lin, Bekey, and Abney, "Robots in War: Issues of Risk and Ethics."
[240] Verdiesen, "Agency Perception and Moral Values Related to Autonomous Weapons"; Marra and Mcneil, "Understanding 'The Loop': Regulating the Next Generation of War Machines."
[241] Worcester, "Autonomous Warfare – A Revolution in Military Affairs."
[242] Giesen and Kristen, "Liability, Responsibility and Accountability: Crossing Borders" p6 ; Kool, "(Crime) Victims' Compensation: The Emergence of Convergence"
[243] See Folkman, "The '8 Great' Accountability Skills For Business Success"; Hoek, van Montfort, and Vermeer, "Enhancing Public Accountability in the Netherlands."
[244] "Accountability."

system development.[245] Australia's system of control "incrementally builds upon itself, embedding controls into military processes and capability at all stages of their design, development, training and usage," and at all stages reviews compliance with national and international legal obligations.[246] The stages of institutional development as presented by Australia, in their most simplified form, are as follows in Figure 7:

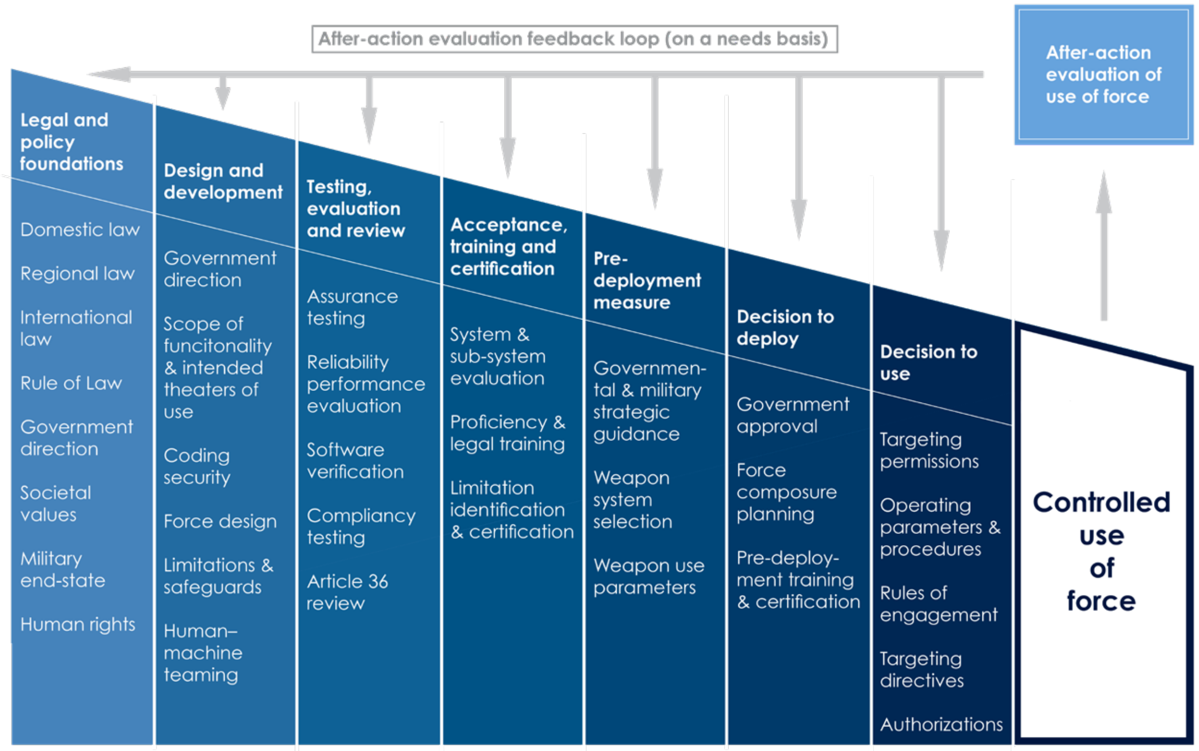


Figure 7. Australia's system of control[247]

Australia also used of this model of a 'system of control' to illustrate its view that the way the phrase 'human control' is often used does not always do justice to, or reckon with, practical military reality. Australia's representatives went so far as to argue that because the phrase 'human control' doesn't adequately cover military reality, it "does not provide a useful basis to further GGE discussions unless there were a common understanding of the term"[248] such as the delineation presented in Australia's model control system in Figure 7.

Despite the fact that the details of this system of control may be up for debate, and although Australia's conclusion that such institutionalized control would solve AWS' responsibility issues is

[245] In the document in question, the term 'control' referred to "the system of processes and procedures through which a state achieves its intended military effect, in a manner compliant with its legal obligations and policy objectives." See "Australia's System of Control and Applications for Autonomous Weapon Systems."
[246] "Australia's System of Control and Applications for Autonomous Weapon Systems."
[247] Adapted from "Australia's System of Control and Applications for Autonomous Weapon Systems."
[248] "Australia's System of Control and Applications for Autonomous Weapon Systems," 5.

premature, this description of a state's development of AWS is very useful to ascertain key points in the processes of accountability. An approach like this can incorporate institutional assessment and constant evaluation of all three intrinsically linked concepts—life cycle, sub-system and OODA loop—which were introduced in Chapter 3 as being crucial to establishing sufficient control to act ethically when using RAS.

An important question, however, remains how far various states will be willing and able to realize such levels of control over private equipment manufacturers and contractors as the system visualized above requires. Another question that arises is what will need to change in this model system of control once there is a shift from the use of algorithms that work based on pre-input criteria towards AI that is more self-learning, like the deep neural networks described in.

5.4 Summary

There are many interpretations of what accountability and responsibility mean exactly. This can be attributed to the difference between usage in practice and on paper, as well as the difference between the people using the term: legal practitioners, policy-makers, private companies and organizations, and wider society may all have slightly differing understandings of what accountability means or should mean. The classic legal frameworks through to establish responsibility for wrongdoing or mistakes are based on the ability to prove individual or group intent behind wrongdoing. However, the fragmentation of military technology development means that direct legal responsibility, accountability and eventually liability are difficult to establish in the diversified and often long life cycles of the elements that make up RAS from R&D all the way past deployment. While other bodies of law have been suggested in order to look at the way to legally address the advent of RAS, there is as of yet no body of law in place that fully suffices. This means that accountability should be addressed at an institutional level all throughout the life cycle.

In light of the increased privatization of military technologies, responsibility is fragmented across many actors. It is therefore crucial to ensure that actors' behavior—be they contractor, military or otherwise—can be accounted for and that ways to ensure and evaluate this are institutionalized in RAS' governance.

6. Conclusion and Recommendations

Robotic and autonomous systems are the latest frontier in the competition for technological dominance in the military domain. While the delegation of tasks to machines is not a new phenomenon, recent advances in computation are enabling machines to carry out increasingly complex tasks. These range from autonomous air-to-air refueling and landing in an independently selected location, to smart swarms of ground, air and naval systems operating in sync through machine-to-machine communication. Alongside technology's fast-paced development, ethical norms evolve continuously. This requires governments and military forces to continuously reflect on how ethical issues and norms of the society they operate in may affect the use of RAS.

This paper has identified four overarching ethical challenges arising from the use of RAS. First, the establishment of human control in increasingly autonomous systems, based on determining *how* and *where* in the life cycle and the observe-orient-decide-act (OODA) loop to maintain control over RAS, as well as *who* should do so over *what* functions of systems. The second challenge is the technical complexity of (semi-)autonomous systems, which leads to decreasing explainability and predictability of the system design, self-learning abilities and software updates. The third challenge is posed by limitations of human cognition, such as automation bias and complacency, as well as the anthropomorphizing of machines, arising in human–machine teaming. Fourth is the institutional risk management of the outsourcing of system design and manufacturing, and RAS interoperability with technologically advanced allied forces.

Considering the increasing proliferation of autonomous systems, including among adversaries, the RNLA should continue to experiment with systems that may enhance its portfolio, without losing sight of foundational ethical principles. In considering the implementation policy for RAS, the RNLA should seek to translate the discussion on meaningful human control into operational terms, such as by identifying controversial or high-risk machine functions, as presented in Chapter 3. This selective approach to establishing and maintaining human control presents a balance between ethical concerns and military objectives. Meanwhile, the debate from Chapter 4 on the extent to which fully autonomous systems could be developed used in ways fully compliant with IHL and could thereby be in line with what society's idea of human dignity may be, is one that underpins the ethical dilemmas surrounding RAS. The break-down of RAS into life cycle, sub-system elements and the OODA loop, as presented in Chapter 3, is relevant again in addressing the challenge that RAS pose for traditional responsibility and accountability in the military, as well as in the broader national and international governance structures in which it operates. The detailed example of a 'system of control' is one of the ways accountability can be thoroughly woven into the institutional handling of RAS the many different stakeholders involved in RAS development. This can help guide decisions on RAS integration into the armed forces in a way that considers ethical standards and ensures moral responsibility at all stages of RAS development and use.

The paper provides the following concrete recommendations to the Dutch government, the Dutch Ministry of Defence and the Royal Netherlands Army, as well as other governments and armed forces at strategic, operational and tactical levels:

Strategic

1. *Institutional development* – adapt internal processes, such as monitoring & evaluation with RAS life cycles, to better address (rapid) technological developments in relation to ethical issues. This means hiring or working closely with experts to guide use case development, testing with private contractors, and facilitating the introduction of RAS within the RNLA;
2. *Ethics by design* – there is a need to develop a set of guidelines for identifying use-cases, designing, validating, and manufacturing ethical RAS in line with the core principles of International Humanitarian Law and Article 36, rather than establishing ethical considerations only at the deployment stage;
3. *Testing* – in determining the appropriate environment for testing RAS, the RNLA may consider studying the testing approaches of other states and determine how to best emulate them while respecting core ethical principles;
4. *Contracting* – identify best practices for military–private sector cooperation in designing, manufacturing, maintaining, and operating military RAS, and delineate legal and moral responsibility for accidents, failures, malfunctions, or misuse of systems among the involved parties;
5. *Transparency* – communicate the Ministry of Defence and RNLA's research into and use of RAS to the public in order to inform and add nuance to the discussion on the value of RAS to the RNLA outside of the dominant 'killer robot' narrative;
6. *Research* – continue to research the role of RAS in the military context, including but not limited to human–machine teaming, embedding of ethics in machines and contingency planning for facing adversarial RAS based on different ethical constructs, and focus on operationalizing these principles into practical applications.

Operational

1. *Human control* – considering the spectrum of autonomy, the RNLA should predetermine where, how and who maintains control over what functions of individual systems, as well as who is responsible for the initiation, use and shut down of systems prior to their deployment;
2. *Selective automation* – identify within what functions and why increasing automation and autonomy will benefit the military without eliciting major ethical concerns, e.g., movement controls, sensory controls and computer vision;
3. *Interoperability* – the Dutch government and the RNLA should push for standardization frameworks among technologically advanced allied forces on the training, deployment and operation of RAS in shared environments or during joint missions;
4. *Design process* – designing of systems should involve the end-users (i.e. operators and supervisors) in the use case development, design and testing phases to ensure the design of human-machine interfaces is suited to those using the systems;
5. *Operation of RAS* – establish and delineate the different levels of freedom within the rules of engagement, based on the degree of autonomy of the system(s) under their command for military units deploying RAS;
6. *Human–machine teaming* – the RNLA should identify the limitations of human cognition in the oversight of RAS, develop understanding of its effect on human-machine teaming, and channel acquired knowledge in the RAS human interface design;

Tactical

1. *Rules of engagement* – within the design and manufacturing process, the RNLA should seek to program fundamental rules of engagement (ROEs) with International Humanitarian Law principles embedded in system design, along with an open architecture to introduce mission-specific ROEs by mission command;
2. *Opacity* – tackle the 'black box' nature of complex systems by developing traceability or logic flow processes to enable operators or supervisors to understand, explain and predict the operation of RAS;
3. *Training manuals* – in cooperation with contractors, issue training manuals for operators, supervisors and commanders of RAS for the initial use and after subsequent software updates that substantially alter the behavior or decision-making process of the system.
4. *Command responsibility* – predetermine command responsibility for the use of RAS for every individual deployment.

Annexes

Appendix A - Expert Session Scenarios[249]

Scenario 1. 'Killerbot'

ISIS has been defeated in Syria by rooting out the last members in Baghouz on the Syrian-Iraqi border. Unfortunately, ISIS ideas have not been eradicated, and an influential leader Abu Bakr-al Baghdadi (ABaB) still prophets his vision of an Islamic Caliphate through his re-occurring presence in the media. ABaB has gone underground, but it is clear that his extreme ideas are gaining traction and groups of young Jihadis are currently being mobilized to plan and conduct terror attacks in their domiciles in Western Europe. It is essential for domestic safety in our country that ABaB is silenced as soon as possible. A major intelligence operation has been conducted and as a result, deployed HUMINT units learned that he is in rural Syria, in a 50km^2 area of mountainous terrain with a myriad of tunnels. The tunnels are not charted and are likely booby-trapped. As such, it will be very difficult to thoroughly comb through the area.

Major Pavlov Strolsky, the local commander of the Russian Spetznatz-unit who was responsible for finding ABaB, has to plan an operation to eliminate ABaB's role as the Jihadi leader. Because of time sensitivity, short exposure time of ABaB, and the need to prevent him from fleeing again, there is no time to discuss alternatives between major Strolsky and his headquarters. It is important to note that communications from within the caves to the outside HQ is impossible due to the iron-ore rich stone of the mountains.

The nations with troops in the area (Russia, Turkey, Syria) are reluctant to conduct searches (Russia and Syria) or are prohibited by their own government (Turkey) through a lack of Rules of Engagement (ROE) concerning these kinds of operations. A solution would be sending out a ground drone loaded with facial recognition software and armed with lethal capabilities. Two months ago, the Russians brought the 'Gusenichnyy' (Crawler) ground drone system into theater for operational evaluation and testing. Civilian (Russian) personnel from Kamaz, the producer of the Gusenichnyy system, assist the military in handling the drone. Although the crawler is a rather small object (60cm high and 35kg), it can move over rocks easily, and is silent, stealthy, and lethal. Experiments

[249] **Disclaimer**: Scenarios 1-4 are entirely fictional but use the names of real companies, states, and locations only for the purpose of a more in-depth simulation. All the names of individuals are fictional. The content of the scenarios is not intended to discredit any companies or states.

in a safe environment have proven that the facial recognition software has 99.95% accuracy and has the ability to self-learn in order to further minimize errors. ABaB's facial features have already been loaded into the 'Gusenichnyy' by civilian engineer Pjotr Pekar of Kamaz. As instructed by the management of Kamaz, Pekar urges Major Stroslky to deploy the crawler for this specific mission in order to get the first real test data on performance, in terms of recognition tasks and lethal tasks.

It is likely that there is only one chance to find and stop ABaB, so if the crawler is deployed, it has to be set on fully autonomous mode. That means that it fires lethally when the software recognizes ABaB. It is Major Strolsky's decision to either send the Gusenichnyy or send his highly trained men into the caves and risk losing them.

Discuss the decision-making considerations, both in ethical and legal sense, which Major Strolsky should or could be contemplating within this case.

Scenario 2. 'Testbed'

The year is 2021 and the Netherlands government is struggling with the last pieces of legislation concerning autonomous armed drones in combat. To fill this gap, it wants to enhance testing, in order to prepare legislation for military use outside of the Netherlands. This knowledge is essential for building legislation relevant to all military activities (including unmanned and lethal) and to show how the military will operate with these systems under water, at the surface or in air and space. Theoretical experiments have been conducted already, but the last piece of information concerning real-life flying and drone-weapon separation (with live weapons) still has to be acquired.

The Commander of the Royal Netherlands Air Force, LtGen Frits Elands, proposes a test over the Vliehors shooting and bombing range on the isle of Vlieland. His plan involves ceasing air operations and training in the area for one day. He proposes the drones take off from Leeuwarden Air Base (which also tested the Reaper MQ9 drone that entered service in 2020) in order to minimize flying time over inhabited areas. The autonomous software is capable of finding the specifically designed target at the range. The drone is programmed to only fire a missile at the predesigned target (which is unique in shape, colors, and pattern). Stringent safety margins, including geo-fencing, are in place.

If and when the test is successful, legislation can be concluded and various unmanned systems owned by the Dutch military will be ready for deployment.

The US urges the Netherlands to take an expeditious approach, as they want more European involvement in their mission in the Sinai (Operation Vanguard against some virulent, widespread, covert operating ISIL units), where a few Dutch troops are currently present alongside the American contingent.

The Netherlands ground drones, developed by the Dutch firm VDL, fill in urgent operational capability gaps in the US operation. VDL is willing to assist in all technical matters to conduct the test, as they foresee great US interest in buying their drone. This would lead to a €2.5 billion deal and 1,500 jobs in the Netherlands for the next 10 years. The pressure on the Dutch government from industry and foreign allies to conduct the testing, finish legislation, and deploy the drones is high.

Discuss the required legal framework and the ethical/moral implications of this testing. What are the consequences for quick deployment of Dutch drones and how relevant is the positive spin-off for the Dutch economy?

Scenario 3. ‘Defend’

Netherlands troops conduct a peace-enforcement operation in the Central African Republic (CAR). Their Camp ‘Hoogeveen’ houses a total of 1,800 Dutch soldiers from the HQ, the Helicopter Task Force, the Special Forces patrol, armored infantry, artillery and support units. No civilians work here, nor is the camp close to built-up areas with local CAR-nationals. The Chief of Defense of the Netherlands, General van de Putte, is currently visiting with the Danish Chief of Defense (General Norrebro), as in a week’s time the Danish will start their operational hand-over from the Netherlands. In one month, the Netherlands will end this mission after four years of fighting local insurgents. Eight Dutch soldiers have died in the past 2.5 years. Intelligence reports indicate that this high-level visit will be used by insurgents to conduct attacks on Camp Hoogeveen as a means of maximum exposure in the media.

Because of these VIP visits, the Danes have provided ground surveillance and protection through their armed drone unit ‘Thor’, which consists of autonomous ground systems that can work in swarms. The Danes have made serious breakthroughs in research concerning drones with high-energy weapons, which are lethal. Because of the terrain, distances, and employed tactics by insurgents, the ground-based drones work on autonomous setting. This means that random search patterns will be ‘walked’ by the drones and their weapons can be fired (based on previously set algorithms and taking the ROEs into full consideration). Unclear to the Dutch staff is what will trigger the drones to use their deadly lasers. Confirmed, however, is that under no circumstance a laser will be used toward or within the boundaries of the camp.

Another means of protection would be stationing almost all soldiers on a rotational schedule outside the Camp’s perimeter. This seems to be of high risk, due to the advanced night-fighting capabilities the insurgents have shown to possess in combination with the level of exhaustion the Dutch troops suffer. The Dutch troops will be rotating out in only a few weeks time after enduring six months of extreme hardship.

It is up to the operational commander of the camp, Colonel Peter van Ellekom to take a decision on defensive measures.

Discuss the decision making, ethics and legality of using a defensive tactic with drones when the 'only' possible victims of an adversary attack are trained Dutch soldiers.

Scenario 4. 'Back to basic'

Netherlands forces are in Lithuania conducting their Enhanced Forward Presence mission (EFP), together with German, English, and Polish troops. They defend the city of Narva (56,350 inhabitants), bordering on the Russian border. Tensions are high, as the Russian S400 system in Kaliningrad has been active for two weeks already and NATO jets performing Baltic Air Policing missions have been 'painted' not only by the surveillance radar, but also by the tracking and fire-control radars of SA7 and SA21 air defense systems. NATO thinks it unwise to deploy manned fighters, due to risks of losing a fighter to a Russian rocket and fears of the escalation that would follow such an attack. In order to de-escalate, NATO deploys some drones instead of armed fighters to keep watch over the ground troops in the city. Western intelligence organizations receive insider information that Russia is planning a devastating air attack on Narva.

In order to defend troops and civilians, defensive measures have to be taken, otherwise a large proportion of city inhabitants and the deployed EFP troops will fall victim to Russia's lethal air attacks within the highly populous medieval town center. NATO's UAVs are not armed (so as not to further escalate), but they are in direct communication with unmanned ground systems equipped with anti-aircraft artillery.

The ground-based drone with these weapons is the so-called 'Umbrella' system. 'Umbrellas' are positioned along the border with Russia and throughout the town on the roofs of high buildings. Due to extremely short reaction times, 'Umbrella' has to be set on autonomous operations. This means that once opposing aircraft will cross the border from Russia to NATO territory, Umbrella will fire upon them. NATO has communicated this to the Russian high command in St. Petersburg and to the Russian President.

Discuss how such a defense strategy would fit within NATO's show of resilience and coherence in defending NATO territory vis-a-vis article 5. Discuss how ethical concerns interact with the prospect of preventing the killing of thousands of civilians.

Appendix B - Expert Session Scenario Summaries

Scenario 1. 'Killerbot'

Killerbot is evidently the most controversial scenario, particularly in terms of human dignity and human control. Trust was a central issue within human control, as concerns were raised over how a system that was only tested in a "safe" (rather than a combat) environment be trusted to complete the mission. Further concern was the inability to communicate with the crawler or abort/

alter the course of action if necessary. Questions arose over meaningful human control (MHC), particularly in terms of what qualifies as a threshold for MHC and whether programming of RAS is enough human control. With regard to human dignity, concern was raised over the inability of the RAS to understand situational nuance, e.g., the target surrendering, use of human shields and presence of non-combatants in the environment. The inaccuracy of 0.05% was seen as a major issue in light of civilians present in the operational environment. As the system had not been tested in a combat environment by the military, the overall performance of the system in line with Article 36 of the Geneva Convention Protocol I was seen as an issue. One group was willing to accept civilian casualties under the Doctrine of Double Effect[250] seeing as the target is a high value target. However, the threshold for the acceptable number of civilian casualties was not determined. As operators of the RAS, responsibility was laid with the Russians (unspecified at what level), but with concerns that the private sector would remain unaccountable for the extent to which it is involved in the development. Moreover, the issue of trust arose, whereby the text was suggestive of the Russian military placing trust in the private sector engineers with little to no experience with the RAS and without regard for Article 36. The self-learning of the system was also perceived as controversial, as this also reduces human control, predictability, and thus, has significant implications for responsibility. Overall, conclusions of whether to send the crawler in were mixed. Some participants said to unequivocally "go for it", while others said the system was doomed to fail and would not meet norms and rules of military engagement. Most participants fell somewhere in the middle, with sharp concerns for human control and human dignity contrasting their recognition that the use of RAS was likely to be the most effective and least risky action given the almost impossible circumstances of the scenario.

Scenario 2. 'Testbed'

The scenario discussions focused primarily on human control and human dignity. The controversy was centered around human dignity, based on the idea that an individual exercise in a controlled environment is not representative of a real combat scenario. As a result, there is insufficient evidence that the UAV would execute 'ethical' decisions in actual combat. Similarly, limited testing highlighted a lack of human control, both in manufacturing and operations. In manufacturing, the principal-agent problem between the private sector and the military resulting from pressure to deploy RAS quickly is likely to result in the deployment of premature systems. In operations, the lack of familiarity of the operator with the UAV could reduce the explainability of the system, in turn reducing operator's control, with implications for responsibility. Human control and responsibility is further at risk due to software patches and self-learning of the RAS, meaning over time explainability would reduce further. As a result, testing will need to be carried out after every software update, as it could alter the system's behavior. More testing is necessary to develop the

[250] The Doctrine of Double Effect "is often invoked to explain the permissibility of an action that causes a serious harm, such as the death of a human being, as a side effect of promoting some good end. According to the principle of double effect, sometimes it is permissible to cause a harm as a side effect (or "double effect") of bringing about a good result even though it would not be permissible to cause such a harm as a means to bringing about the same good end." See McIntyre, "Doctrine of Double Effect."

operator's familiarity with the system to limit the possibility for the operator to be sheltered from responsibility by claiming that the machine acted unpredictably. As a result, one group argued that testing should focus on confirming operational parameters and the functioning of the system, rather than simulating operational behavior, which would still be far from an actual operational environment. Furthermore, test environments are not capable of testing the chain of command, and hence responsibility, as the outcomes are predetermined and hence, the operators do not experience the pressure of a real life scenario.

Solutions proposed to the challenges were based on an understanding that further testing would be necessary, but the pursuit of RAS will continue, driven by competition from adversaries and the faster pace of technological development in the private sector. One group suggested that testing can also provide insight into the ways other actors will use RAS, even if the Dutch government does not allow the deployment of certain RAS in the future. One solution was to focus on 'development' legislation rather than 'deployment' legislation, arguing that this would enable further testing and more extensive control over the private sector. One of the groups also noted the difficulty of legislating unpredictable behavior. In the case that testing is limited, RAS should only be allowed to carry out missions highly comparable to those it has executed in a controlled environment.

Scenario 3. 'Defend'

With the scenario as it was written, there was a consensus that the lethal autonomous system could not be used. This was largely down to the fact that the Netherlands troops were not familiar with the Danish system in question and they did not know what would trigger the lethal ray the system's swarmed robots could shoot. All groups concluded that for the use of this, or any autonomous system in the military, it is crucial that the troops using it fully understand the system they are using and can predict the outcome of whatever situation or environment the system is used in. What the necessary level of (training) experience with a system is, or how much information the commander needs to receive from a party delivering the system remained a point of discussion, and is clearly a topic that requires further research. Some participants were of the opinion that Dutch troops would need to be well acquainted with and have trained themselves for quite some years with RAS before being able to ethically and predictably deploy them. Others thought that when collaborating with trusted partners it could be enough to receive a certain amount of information on the systems used, including perhaps testing reports, relevant indicators, ROEs, and more such data. Difference in levels of not only confidence but also trust in fellow soldiers as opposed to RAS is partly down to the extent to which one can understand the reasoning process of each. Even if a human acts irrationally, there is a certain reasoning process behind the actions that can be explained afterward. As close as possible a level of understanding is needed of the reasoning process that guides the actions of RAS.

Another question that came to the foreground was whether the full level of autonomy was even necessary in this situation. Many found that, if given the option, the best solution would be the teaming of a semi-autonomous swarm and human response in case of threat or confrontation. Reasons to reconsider this unwillingness to opt for full autonomy came in the form of situations of pressure, such as where there is severe time pressure or an environment in which humans could perform well enough, e.g., difficult terrain or severe exhaustion. Choices are made on the basis of (imminent) risk analysis and the question of how necessary a certain level of autonomy is to prevent the further endangering of human lives. Even so, it was said that such systems should have the option to perform in various modes and not as autonomous systems only, and that there should be serious consideration of in which cases or settings lethality is an option.

Some groups already went in the direction of solutions, with the most concrete being the development of an international, standardized system for the various types of RAS that partners may use. Once the use of RAS for various purposes in international contexts is more normalized, it will be important that partners can be quickly made aware of which system it is they are dealing with, what this type of system's outcomes are based on, how it has been tested, how the system can be used, and more such crucial aspects of confidence in decisions on the (joint) use or avoidance of RAS in certain military contexts.

Scenario 4. 'Back to Basic'

Scenario 4 was generally straightforward and uncontroversial. There was an overall consensus among the groups that the umbrella system should be deployed, with one group going as far as to state that "it would be unethical not to deploy the RAS". The main reason for the conviction was that the system is defensive by nature and that the actions were explicitly communicated to the Russian higher command. The system is likely to serve as a better deterrence tool as it is in autonomous mode, thus limiting the strategic choices of the adversary. Some noted that it would be important for the adversary to know the demonstrated potential of the RAS to limit the temptation to test the system. Ethical issues were almost untouched, with groups rather side-tracking to operational and political issues, notably a point flagged by one of the groups that the Russians may perceive the deployment of RAS on the border as an escalation. Frequent comparisons were drawn to the Aegis system, the Patriot missiles and the Israeli Iron Dome, so the 'availability heuristic' was quite prevalent across the groups. Primary controversies in ethics were false positives, i.e. the shooting down of non-military aircraft (e.g. USS Vincennes incident) or situations like the Russia-Turkey dispute over the shooting down of the Russian aircraft over the Turkey-Syria border, with the idea that rapid decision-making by the RAS can lead to unnecessary escalation if the threat is not explicitly demonstrated. Particularly due to the proximity of the border, the area of the umbrella system's operation should be clearly defined to avoid takedowns of aircraft still within Russian airspace. Some disagreements were evident in terms of human control, with arguments between autonomous mode vs keeping humans in the loop. Little discussion on NATO Article 5, with most agreeing that Article 4 was more fitting for the scenario, with the possibility of an escalation to Article 5 should an attack occur. To address proportionality concerns, one group suggested illuminating/marking a target first and firing only in case of violation of warnings.

Appendix C - List of Abbreviations

ACTUV – Anti-Submarine Warfare Continuous Trail Unmanned Vessel

AGI – Artificial General Intelligence

AI – Artificial Intelligence

ASW – Anti-submarine warfare

(L)AWS – (Lethal) Autonomous Weapon Systems

CCW – Convention on Certain Conventional Weapons

DL – Deep learning

DNN – Deep neural network

EFP – Enhanced Forward Presence

GGE – Group of Governmental Experts

HARM – High-speed anti-radiation missile

HCSS – *The Hague* Centre for Strategic Studies

ICJ – International Court of Justice

ICL – International Criminal Law

ICRC – International Committee of the Red Cross

IHL – International Humanitarian Law

ILC – International Law Commission

LOAC – Law of Armed Conflict

MHC – Meaningful human control

ML – Machine learning

NATO – North Atlantic Treaty Organization

OODA – Observe-orient-decide-act

PMSC – Private military and security contractor

R&D – Research and development

RAS – Robotic and autonomous system

RNLA – Royal Netherlands Army

ROE – Rules of engagement

TNO – Netherlands Organisation for Applied Scientific Research

UAV – Unmanned aerial vehicle

UGV – Unmanned ground vehicle

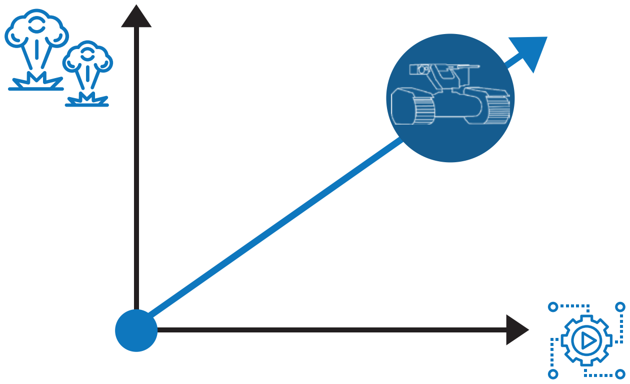
Factsheet

Managing RAS: The Need for New Norms and Arms Control

Regulating Robotic and Autonomic Systems

Should controversial areas of RAS, such as (lethal) autonomous weapon systems ((L)AWS), be regulated? If so, how?

Not all RAS are Created Equal



Not all RAS are equal. Mainly the systems with a high degree of autonomy in combination with a 'use-of-force'-function are regarded as controversial. The debate on governing such systems is gaining momentum.

The Status Quo

Currently, debate on autonomous weapons takes place in the Group of Governmental Experts (GGE) within the UN. This effort alone isn't sufficient - other fora are necessary to generate broad consensus on the regulation of autonomous weapons.

Challenges for Regulations

There's several challenges that may make regulations less useful and/or effective:

- New technological developments out-pacing lawmakers.
- Arms control, like with other weapon systems, is - currently - too ambitious.
- The private sector is in the lead when it comes to technological innovation.

Potential Paths Forward

Hard law often involves parliamentary ratification procedures and clarity: regulation of some (L)AWS would become binding law

Soft law is the most realistic option. It fits in the broad development towards 'principles-based' policies

Stakeholder involvement & Trusted Communities

The involvement of stakeholders, such as industry, is pivotal for governments to be able to keep up with technological developments, empowering them in negotiations on regulation.

Trusted communities enhance understanding between government and business, also on the topic of regulation.

Route Markers for Future Regulation

- The effects of high- and full-level autonomous weapons with defensive or offensive functions are to be (additionally) regulated
- This debate is part of a broader debate on technology and human-machine interaction
- Alternative fora might be needed to establish broad consensus on RAS-regulations
- There's a preference for softer/voluntary law/ instruments for regulation
- Development of 'trusted communities' offers a promising way forward for the Netherlands

The Hague Centre
for Strategic Studies
Lange Voorhout 1
2514 EA The Hague
The Netherlands
hcss.nl

The Hague Centre for Strategic Studies

Chapter 3

Clingendael
Netherlands Institute of International Relations

Managing RAS: The Need for New Norms and Arms Control

Hugo Klijn and Maaike Okano-Heijmans (Clingendael Institute)
with contributions from Bianca Torossian (HCSS)

Executive Summary

Against a backdrop of geopolitical tensions and rapid technological developments, the debate on governing autonomous weapons is gaining momentum – both in the Netherlands as well as in other countries. Discussions are complicated, however, because of wide variations in the positions of countries, compounded by a tendency of some politicians and non-governmental organizations to frame the discussion in alarmist terms.

The regulation of controversial categories of robotic and autonomous systems (RAS) requires new approaches and new instruments. Building on theories of transnational governance, this paper highlights so-called trusted communities as a potentially valuable instrument to engage relevant stakeholders, particularly those from the private sector. Apart from continuing its efforts in formal frameworks – such as the Convention on Certain Conventional Weapons (CCW), where Lethal Autonomous Weapon Systems (LAWS) are discussed – the Netherlands government may consider reaching out to businesses and relevant experts at home as well as in like-minded states through trusted communities. Such networks have the ability to bring together key actors to provide input for developing principles and norms for further regulation and export control regimes that are based on mutual trust and respect.

1. Introduction

As robotic and autonomous systems (RAS) can perform increasingly advanced functions, the debate on governing autonomous weapons is gaining momentum. International positions still differ widely, ranging from proponents and opponents of a ban on such weapons to a group of countries that lie in between and emphasize the need for further clarification and elaboration of existing regimes. For their part, civil society and some politicians often tend to frame the discussion in alarmist terms, speaking of 'killer robots'. More broadly, however, the debate involves technological possibilities, prospects of military applications, ethical questions, and the need for control or regulation.

This paper addresses the latter issue and puts forward thoughts regarding the regulation of controversial areas of Robotic and Autonomous Systems (RAS) – more specifically, of (lethal) autonomous weapon systems (AWS or LAWS).[1] Key questions to address include why (L)AWS regulation is (not) needed, what forms it may take, and how to organize this. (L)AWS regulation should concern the Netherlands as a country traditionally supportive of the rules of international law and multilateral agreements in general. Also, the Dutch government has an interest in steering the intensifying domestic debate on the use of autonomous weapons systems in armed conflict, both inside and outside parliament, and needs to deliver on new thinking and action.[2]

In order to gain clarity on the regulation of controversial areas of RAS, chapter 2 considers prevailing purposes of arms control and asks whether at this stage gray zones may already be identified. Chapter 3 looks at the current state of regulation and briefly discuss international positions with regard to autonomous weapons. In chapter 4 the challenging environment in which the debate takes place is described, including elaborations on technology, arms control and the role

[1] The authors of this paper gratefully acknowledge the insights gained at an expert session held on 13 November 2019 at the Clingendael Institute in The Hague. The stated goal of this particular session was "*to identify and assess the international community's available options in managing RAS, and the technological, geopolitical, and legal feasibility of developing new norms and functioning arms control arrangements in this area*". Bianca Torossian et al., "The Military Applicability of Robotic and Autonomous Systems," Security, HCSS Security (The Hague: The Hague Centre For Strategic Studies (HCSS), March 1, 2019). This paper and the expert session are part of an overall project carried out by the Hague Centre for Strategic Studies (HCSS) concerning "The Military Application of Robotic and Autonomous Weapons (RAS): What, Why, How and Under What Conditions?", commissioned by the Royal Netherlands Army. This paper is published separately from the forthcoming capstone document on the project. Responsibility for the content and for the opinions expressed rests solely with the authors; publication does not constitute an endorsement by the Netherlands Ministry of Defence.

[2] In the Netherlands, the Advisory Committee on Issues of International Law (CAVV) of the Advisory Council on International Affairs (AIV) addressed these topics in an advice on autonomous weapon systems, published in October 2015. In its response, the Dutch Cabinet subscribed to the key finding of this report, which holds that meaningful human control is required for the use of autonomous weapon systems. Considering the rapid developments in robotics and AI and the evolving international debate, the usefulness of this advice is again considered in 2020. See Netherlands Advisory Council on International Affairs, "Autonomous Weapon Systems: The Need for Meaningful Human Control" (Netherlands Advisory Council on International Affairs, October 2015), https://aiv-advies.nl/8gr#government-responses.

of the private sector as well as the shifting balance between (i) the defense industry and (ii) civilian research and development spheres. Chapter 5 deals with potential paths forward in terms of hard, soft, and voluntary instruments. This section also discusses stakeholder involvement and proposes a refocus from a rules-based to a principles-based approach by way of voluntary instruments as a promising way to address the uncertainties of a system in flux.

Such voluntary instruments may be less than ideal in the eyes of certain policymakers, civil society actors, and private actors who would generally prefer more formal instruments that have been common in arms control. Undeniably, moving away from established practices that were successful in the past is difficult. But it needs to be acknowledged that the current geopolitical climate – characterized by great power rivalry and a diminishing commitment to a multilateral rules-based system – seems hardly conducive to new multilateral arms control agreements. The propensity to rely on hard-law instruments may be reconsidered in favor of more innovative thinking also on soft and voluntary instruments – that is, new approaches to regulating controversial areas of RAS.

Building on theories of transnational governance, this paper highlights trusted communities as a potentially valuable instrument to engage relevant stakeholders, particularly those from the private sector. It is suggested that in addition to continuing its efforts in formal frameworks – such as the Convention on Certain Conventional Weapons (CCW), [3] where (L)AWS are discussed, or the Wassenaar Arrangement on transfers of conventional arms and dual-use goods and technologies – the Netherlands government may consider reaching out to businesses and relevant experts at home as well as in like-minded states through so-called trusted communities. Such networks have the ability to bring together key actors to provide input for developing principles and norms for further regulation and export control regimes that are based on mutual trust and respect.

[3] In full: Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects. United Nations Group of Governmental Experts, "Draft Report of the 2019 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems" (Geneva: United Nations GGE, August 21, 2019), https://www.unog.ch/80256EDD006B8954/(httpAssets)/5497DF9B01E5D9CFC125845E00308E44/$file/CCW_GGE.1_2019_CRP.1_Rev2.pdf.

2. Regulating RAS: why and what?

Experts hold different positions on the issue of what purposes of regulating autonomous weapons should prevail. Although there is a school of thought stating that regulation should only be factored in when human control is absent, there seems to be a growing consensus on the need for (some form of) regulation under the current circumstances, or at least at a stage that precedes 'full autonomy'. In this context, moral purposes are sometimes advanced ('leading by example'), while at other instances there is a more general call for a higher level of transparency (in order to increase predictability) and arrangements to avoid proliferation to 'bad' or non-state actors.

As autonomous weapons are likely to increase the speed of armed conflict and make reaction times much shorter, taking humans 'out of the loop' may become an incentive *per se* to accelerate the development of autonomous weapons. In that context, curbing such 'first-mover advantage' can also be a motivation for regulation.[4] Otherwise, attention is being given to ensuring the safety and reliability of systems through regulation and/or standardization (given possible proneness to spoofing or hacking). Other suggestions point in the direction of focusing on 'white-listing' during this early phase of development: drawing up principles that outline what uses are allowed, rather than what is forbidden. This latter idea is also informed by the fact that even if countries favor 'bans', it is not clear what exactly should be banned.

Most, but not all, purposes mentioned in the context of regulating autonomous weapons are not alien to the underpinnings of existing arms control frameworks. Still, taking into account the forward-looking nature of this debate and the many unknowns in this respect, there is a specific focus on generating more transparency and opting for positively framed recommendations instead of more classical prohibitive measures.

Current rules, standards, and practices are relevant but, most probably, insufficient to cover developments with regard to autonomous weapons.[5] At the very least, these developments would require refinements of existing regulation. Among other things, the central notion of 'human control' needs continuous further elaboration, and perhaps at some point a 'threshold' should be identified in this respect determining whether technologies are subject to existing regulation or

[4] An intensifying first-mover advantage would create an incentive "to develop AWS first and ask strategic questions later": Nathan Leys, "Autonomous Weapon Systems and International Crises," *Strategic Studies Quarterly*, no. Spring 2018 (2018): 51.

[5] A parallel can be drawn to the regulation of the company Uber, which claims it is not a taxi service and therefore does not have to comply with taxi regulation. One may also think of the regulation of Facebook vis-à-vis media and publishing rules.

not. Furthermore, the limited verifiability of processes, such as supply chains, is a point of concern and calls for high levels of built-in trust: something to be considered in regulation efforts.

Given the ‘moving target’ nature of autonomy in weapon systems, it is difficult to establish where exactly gray zones, or even blank spots, occur in the current regulatory landscape. Still, a broadly shared point of view is that, no matter what, more regulatory progress should be made. For the time being, this will amount to further discussion, formulating additional general principles, and trying to refine existing rules in order to gear them toward future technological developments. The further refinement of ‘human control’ remains an important element in this endeavor.

It is commonly stated that technology itself is neutral and can never be illegal or immoral. Technology always offers both risks and opportunities. It is therefore important not to frame technology as a looming threat, or as something to be curbed or to be pitted against society. The *use* or *application* of technology may very well be illegal or immoral. Accordingly, for the purpose of this paper, it is the ‘usage’ and ‘effects’ of applied technologies (in this case when applied militarily) that are of interest.

It is important to emphasize that this analysis focuses on specific subsets of the broad scheme of RAS – namely (L)AWS combined with a broad degree of autonomy, which, taken together, make for controversy. Figure 1 schematically presents various categorizations of RAS, distinguished in four pillars: service and support; information and intelligence; (self-)defensive use of force; and offensive use of force.

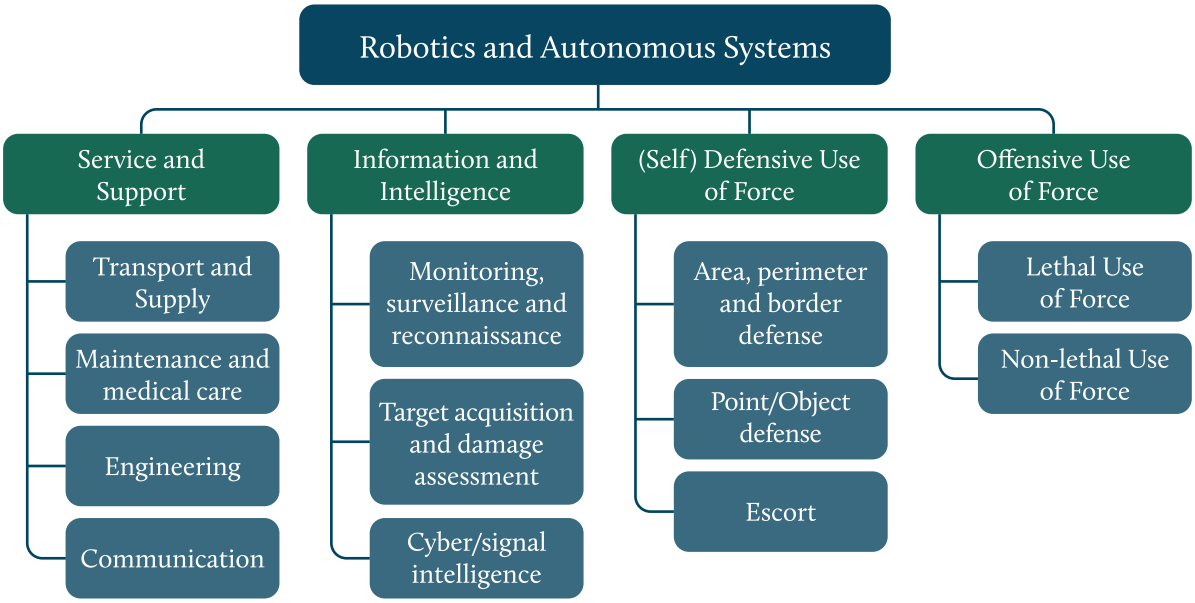
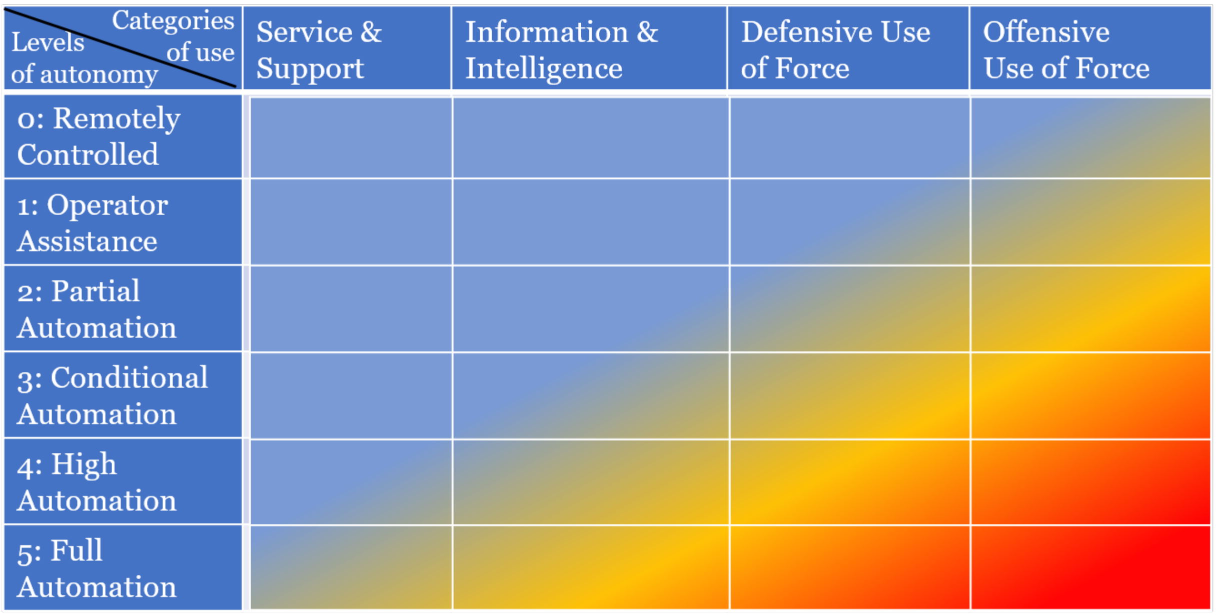

Robotics and Autonomous Systems
Service and Support
Transport and Supply
Maintenance and medical care
Engineering
Communication
Information and Intelligence
Monitoring, surveillance and reconnaissance
Target acquisition and damage assessment
Cyber/signal intelligence
(Self) Defensive Use of Force
Area, perimeter and border defense
Point/Object defense
Escort
Offensive Use of Force
Lethal Use of Force
Non-lethal Use of Force

Figure 1. Categorization of RAS.[6]

[6] Torossian et al., “The Military Applicability of Robotic and Autonomous Systems.”

In order to establish what constitutes the controversial areas of regulating RAS, these four categories may be linked to the six levels of automation commonly distinguished. As illustrated in Figure 2, these six levels are remotely controlled systems; operator assistance; partial automation; conditional automation; high automation; and full automation.[7]

When creating a cross section of these two schemes ('categories of use' and 'levels of autonomy') it becomes apparent that not all RAS are controversial. For example, an escort system (under defensive use of force) that is partially automated or a transport and supply system (under service and support) that is fully automated, are not controversial, and hence not discussed in the context of arms control or this paper. Controversial areas include the cross section between high- and full-level automation and defensive and offensive use-of-force functions (see Figure 2). In Figure 2, the red-to-blue gradient (whereby red denotes the highest degree of controversy and blue denotes the lowest degree of controversy) shows the degree of controversiality of systems that fall in each cross-section of RAS use-type and level of autonomy. It is intended as general depiction only.



Categories of use Levels of autonomy	Service & Support	Information & Intelligence	Defensive Use of Force	Offensive Use of Force
0: Remotely Controlled				
1: Operator Assistance				
2: Partial Automation				
3: Conditional Automation				
4: High Automation				
5: Full Automation				

Figure 2. Levels of automation and subsets of automated systems.[8]

[7] In doing so, the relevant level of autonomy needs to be assessed at four different categories of performance, namely execution of the core task; monitoring the environment; fallback performance; and performance modes. Torossian et al.

[8] Authors' compilation.

3. The current state of regulation

In order to identify possible next steps in the field of regulating (L)AWS – both in attention and in instruments – this section discusses mechanisms that are currently in place. In addition, we address the question of whether existing mechanisms of arms control are applicable to (L)AWS (and vice versa) and what mechanisms may still be explored.

3.1 The status quo

The most substantial multilateral debate on autonomous weapons takes place within the framework of the CCW, in particular in the Group of Governmental Experts (GGE), which includes High Contracting Parties and Signatory States to the Convention, some States outside the Convention, and representatives from international organizations, non-governmental organizations, and academia.[9] One of the GGE's tasks is to consider "[p]ossible options for addressing the humanitarian and international security challenges posed by emerging technologies in the area of lethal autonomous weapons systems in the context of the objectives and purposes of the Convention without prejudging policy outcomes and taking into account past, present and future proposals."[10] The GGE has formulated a set of informal guiding principles that have been adopted by the CCW, the latest entry of which is:

> *"Human-machine interaction, which may take various forms and be implemented at various stages of the life cycle of a weapon, should ensure that the potential use of weapons systems based on emerging technologies in the area of lethal autonomous weapons systems is in compliance with applicable international law, in particular IHL. In determining the quality and extent of human-machine interaction, a range of factors should be considered including the operational context, and the characteristics and capabilities of the weapons system as a whole."*

Although it is encouraging to learn that the CCW/GGE has been able to come up with guiding principles, it took several years of discussion to draw up a list of guidelines that are still very general in nature. Due to the wide variety of positions in this large group, it should not come as a surprise

[9] For a GGE list of participants and other related documents, see The United Nations, "2019 Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)," United Nations Geneva, n.d., https://www.unog.ch/80256EE600585943/(httpPages)/5535B644C2AE8F28C1258433002BBF14.

[10] See the GGE 2019 report: "Draft Report of the 2019 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems."

that to some extent these talks were encumbered and slow.[11] While the consensus-based CCW/GGE still counts as a necessary tool to further this debate, it is doubtful whether this effort alone is sufficient.[12] Despite the inclusion of NGOs and academia, state parties are dominant in this format, and industry is only present in a backbench capacity. The question is warranted, therefore, whether there is a need for other platforms next to the CCW/GGE to contribute fresh thinking to this topic in order to move from so-called 'thin state consent' to 'thick stakeholder consensus'.[13] Indeed, a growing group of experts seems to be of the opinion that innovative approaches are needed to share (technological) information intelligently and to push forward the debates on definitions, norms, and standards. Countries like the Netherlands need to decide on the directions of the modernization of their armed forces and their international posture amid an intensifying public debate. Consequently, the Dutch government should further develop its approach in order to deal with rapid technological developments in a changing international context.

A challenge complicating the debate on regulating autonomous weapons is the so-called Collingridge dilemma.[14] This dilemma holds that when trying to control technology, society first suffers from a lack of information about the technology's impact, and once the technology has become entrenched, society then lacks the power to control it. As a rule, regulation, especially when multilateral, will trail behind developments. This is probably even more salient in the case of RAS, as the private rather than the public sector is leading in the design and development of relevant technologies, and the latter has to bridge a knowledge gap before embarking on regulation.[15]

Existing regulation relating to autonomous weapons revolves primarily around International Humanitarian Law (IHL), while references to Human Rights Law are much more controversial and have not been met with consensus in multilateral fora. IHL, the corpus of 'laws of warfare', contains provisions about the principles of distinction, proportionality, and precautions that govern the employment of weapon systems to mitigate the effects of armed conflict. In this context, a specific regulation is formed by Article 36 of the Additional Protocol I to the Geneva Conventions, which requires states to subject *any* new weapon to legal review, ensuring that the abovementioned principles are respected. One might argue this constitutes a sufficiently binding framework to

[11] According to a CCW/GGE participant.

[12] Even critical voices maintain that the CCW has clarifying power and serves as a catalyst (Neil C. Renic, "Death of Efforts to Regulate Autonomous Weapons Has Been Greatly Exaggerated," *Bulletin of the Atomic Scientists*, December 18, 2019, https://thebulletin.org/2019/12/death-of-efforts-to-regulate-autonomous-weapons-has-been-greatly-exaggerated/.). Meanwhile, the UN Under-Secretary-General of Disarmament Affairs has stated she believes the CCW could still agree on key measures in the run-up to its December 2021 review conference (Janosch Delcker and Andrew Gray, "Top UN Official: It's Not Too Late to Curb AI-Powered Weapons," *POLITICO*, February 13, 2020, https://www.politico.eu/article/top-un-official-its-not-too-late-to-curb-ai-powered-weapons/.)

[13] Terms introduced by legal scholars Pauwelyn, Wessel and Wouters in a 2012 working paper on the stagnation of international law (Joost Pauwelyn, Ramses A. Wessel, and Jan Wouters, "The Stagnation of International Law," Working paper (Leuven: Leuven Centre for Global Governance Studies, 2012), https://research.utwente.nl/en/publications/the-stagnation-of-international-law.). This paper follows this approach in its discussion of 'trusted communities', below.

[14] Point made by Maaike Verbruggen of the Vrije Universiteit Brussel at the expert session on 13 November 2019, the Clingendael Institute.

[15] In this respect, lessons may already be learnt from other AI applications, where initial laissez-faire policies toward industry have led to later calls for technology bans and post-effect regulation. See Mark MacCarthy, "AI Needs More Regulation, Not Less," *Brookings* (blog), March 9, 2020, https://www.brookings.edu/research/ai-needs-more-regulation-not-less/.

cover autonomous weapons, but it is a public secret that only a handful of states actually adhere to this provision and that the process is in fact anything but transparent.

This raises the question of whether, for instance, an extra protocol on autonomous weapons should be added to the Convention on Certain Conventional Weapons (CCW) or whether the use of these weapons may be (partly) regulated elsewhere in the 'vast and pillarized' arms control architecture.[16] Again, the elusiveness of autonomous functionalities and the blurring of traditional lines between munitions, platforms, and/or delivery systems complicate matters. Within the CCW, as the most prominent diplomatic venue where autonomous weapons are being discussed, positions still vary widely: one end of the spectrum maintains that 'autonomy' is covered by existing regulation and requires no new approach, whilst on the contrary, those at the other side of the spectrum believe that a total ban should be initiated.

Finally, with regard to regulation of autonomous weapons systems, various initiatives have been launched outside the traditional arms control community. This is not a unique phenomenon (one may think of earlier and ongoing NGO and scientists' campaigns concerning weapons of mass destruction or conventional devices),[17] but certainly one applying to RAS, with such initiatives now emanating from the private tech industry as well. Therefore, (future) RAS regulation must reckon with a wider stakeholder community, which will be both a challenge and an opportunity to broaden a support base for further decision-making.

3.2 Conceptualizing control mechanisms

In this paper, the various potential control mechanisms of (L)AWS are divided into three categories, namely hard law, soft law, and voluntary measures. Hard law concerns binding treaties that are negotiated and agreed upon between states. For its part, soft law involves quasi-legal instruments such as politically binding Codes of Conduct (CoCs) or Confidence and Security Building Measures (CSBMs), sometimes involving multiple stakeholders other than states. Finally, voluntary instruments include behavioral principles or norms and exchanges of best practices or other information, within or outside traditional arms control communities. These may be developed within so-called trusted communities (as elaborated upon below) that aim to further information sharing between the public and private sector, and thereby to build confidence and encourage restraint. The three categorizations are illustrated schematically in Figure 3.

[16] For a catalog of treaties and agreements drawn up by the US Congressional Research Service in 2019, see Amy F. Woolf, Mary Beth D. Nikitin, and Paul K. Kerr, "Arms Control and Nonproliferation: A Catalog of Treaties and Agreements" (Congressional Research Service, March 18, 2019), https://crsreports.congress.gov/product/pdf/RL/RL33865.

[17] Notable cases are the Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on their Destruction ('The Ottawa Treaty') and the Convention on the Prohibition of the Development, Production, Stockpiling and the Use of Chemical Weapons and on their Destruction ('The Chemical Weapons Convention').

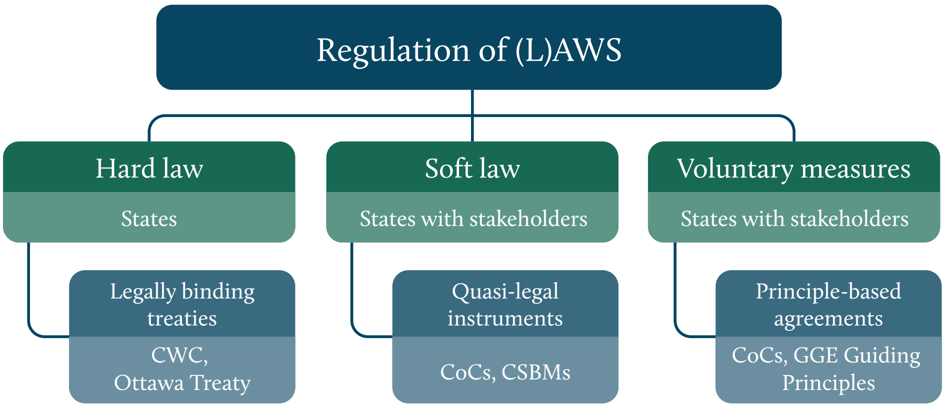


Figure 3. The three mechanisms for regulating (L)AWS.[18]

3.3 Diverging approaches on the multilateral level

States hold varying positions on how regulatory mechanisms should apply to (L)AWS. Broadly speaking, three groups of countries can be distinguished. The first is that of countries which are explicitly in favor of banning (L)AWS. This group includes a number of Latin American, African, and Asian countries. Some European countries, such as Belgium and Austria, have joined their ranks. This group is at odds with countries that have expressed themselves against a ban, such as the US, Russia, and Israel.[19] A third, heterogenous group lingers in between and largely subscribes to the need for further clarification and elaboration of existing regimes. Within Europe, Germany represents this latter school of thought and, supported by countries like Sweden and the Netherlands, is actively engaged in discussing these matters.[20]

Germany, together with France, has also spoken out in favor of a political declaration on autonomous weapons with an eye to further elaborating principles regarding human control and accountability. At the same time, it should be noted that France's position is slightly ambiguous in that it refers to 'fully' autonomous weapons, a position that would not cover a whole range of systems. China's position in this regard has been characterized as 'strategic ambiguity', combining a 'restriction through law' perspective (nominally it belongs to the 'ban group', but only regarding the *use* of (L) AWS) with the active pursuit of AI-enabled military applications.[21]

[18] Authors' compilation. Codes of Conduct (CoCs) are categorized under both 'Soft law' and 'Voluntary measures' as different CoCs bind parties in different ways. For example, the OSCE Code of Conduct on Politico-Military Aspects of Security is 'politically binding' for all participating States of the OSCE, whereas the Hague Code of Conduct against Ballistic Missile Proliferation is a voluntary non-binding instrument open to all states. These two principle-based agreements thus highlight the thin line between soft law and voluntary measures.

[19] Hayley Evans and Natalie Salmanowitz, "Lethal Autonomous Weapons Systems: Recent Developments," *Lawfare* (blog), March 7, 2019, https://www.lawfareblog.com/lethal-autonomous-weapons-systems-recent-developments.

[20] In March 2019, the German Foreign Ministry organized the conference 'Capturing Technology: Rethinking Arms Control'. Details available at: Douglas Barrie et al., "2019. Capturing Technology. Rethinking Arms Control. Conference Reader" (German Federal Foreign Office, March 15, 2019), https://rethinkingarmscontrol.de/wp-content/uploads/2019/03/2019.-Capturing-Technology.Rethinking-Arms-Control_-Conference-Reader.pdf.

[21] Elsa Kania, "China's Strategic Ambiguity and Shifting Approach to Lethal Autonomous Weapons Systems," *Lawfare* (blog), April 17, 2018, https://www.lawfareblog.com/chinas-strategic-ambiguity-and-shifting-approach-lethal-autonomous-weapons-systems.

4. Challenges to regulatory usefulness and effectiveness

A number of unique characteristics inherent to this discussion make it challenging to debate what form of arms control may be effective in managing controversial categories of RAS. The first relates to technology, the second to regulation or arms control, and the third to the role of the private sector.

4.1 New (uses of) technology

Whether the conversation is about RAS, (L)AWS or Artificial Intelligence (AI) in military affairs, the application of new *technology* is the common thread. These technologies are developing fast and offer sometimes spectacular prospects for their use in both military and civilian applications, hence the temptation of some politicians and NGOs to hone in on alarmist scenarios and frame the discussion exclusively in terms of 'killer robots',[22] or 'drone swarms'.[23] Taking a step back, however, one should realize that this topic is part of a larger technology debate that has only recently been gaining traction. This discussion arguably differs from earlier technology debates because of the aforementioned pace and qualitative leap of developments. Autonomy within weapon systems has been around for quite some time, but the increasing relevance of human-machine interaction and the prospects of 'machine learning' enabled by AI heighten this matter to a new level.

Another feature that distinguishes today's discussion is its cross-sector character: it is waged in the civilian sphere no less than in the military sphere, since the technologies concerned originate from various sources and are widely applicable. There are similarities between the arguments over facial recognition technologies for surveillance purposes and arguments over autonomous functionalities for military purposes.

Finally, these technologies promise to be relatively cheap, easily accessible, and "almost invisible except when [they blink] off."[24] The fact that autonomy is not a static function of weaponry further

[22] See for instance: "The Campaign To Stop Killer Robots," The Campaign to Stop Killer Robots, 2018, https://www.stopkillerrobots.org/. "The Campaign To Stop Killer Robots."

[23] See for instance: Zachary Kallenborn and Philipp C. Bleek, "Drones of Mass Destruction: Drone Swarms and the Future of Nuclear, Chemical, and Biological Weapons," *War on the Rocks* (blog), February 14, 2019, https://warontherocks.com/2019/02/drones-of-mass-destruction-drone-swarms-and-the-future-of-nuclear-chemical-and-biological-weapons/.

[24] Kevin Kelly, "The Three Breakthroughs That Have Finally Unleashed AI on the World," *Wired*, October 27, 2014, https://www.wired.com/2014/10/future-of-artificial-intelligence/.

complicates the matter, because this elusiveness means that the discussion is not always about identifiable (weapon) systems, as is the case with existing regimes that mostly regard specific categories such as chemical, biological, or nuclear weapons, or certain types of conventional arms or delivery systems.

Although the discussion focuses on autonomous weapons, the possible applications of artificial intelligence in the military domain in a wider sense concern the "optimization of automated processing (e.g. improving signal-to-noise ratio in detection), decision aids (e.g. helping humans to make sense of complex or vast amounts of data), and autonomy (e.g. a system taking actions when certain conditions are met)."[25] At the same time, one may also want to take into consideration that, until now, machine learning developments have covered a relatively circumscribed field, in that they have resulted in capabilities "more efficient at solving existing tasks rather than tapping into new tasks on their own."[26] Similarly, the 2016 Stanford University One Hundred Year Study on Artificial Intelligence found that so-called 'general AI' (able to make decisions on its own) is not likely to be developed in the near future.[27] This suggests that there is still a significant gap between 'narrow AI' (which is described as problem-solving tools designed to perform specific narrow tasks, which have already existed for some time) and 'general AI' that involves technologies mimicking and recreating functions of the human brain, which has a long way to go.[28] These observations are not meant to disregard potentialities (it is a given that technological possibilities will always shape the struggle for advantage) but rather to demystify part of the ongoing discourse.

4.2 Arms control

From the outset, it should be recognized that in the context of robotic and autonomous systems at this stage, 'arms control' is a highly ambitious and perhaps somewhat premature goal. Given the early and complex phase of the RAS debate, and the lack of common language on definitions and categorizations, the term 'arms control' sets the bar rather high compared to existing arms control regimes which relate to established, well-defined weapon categories. Furthermore, the current geopolitical climate seems hardly conducive to new multilateral arms control agreements in the first place. Apart from the difficulty of engaging the likes of Russia and China, the United States too has been retreating from or undermining a multitude of old and newer multilateral agreements in various domains, including the Paris Climate Agreement, the Intermediate-range Nuclear Forces treaty (INF) and the World Trade Organization.

[25] Larry Lewis, "Killer Robots Reconsidered: Could AI Weapons Actually Cut Collateral Damage?," *Bulletin of the Atomic Scientists* (blog), January 10, 2020, https://thebulletin.org/2020/01/killer-robots-reconsidered-could-ai-weapons-actually-cut-collateral-damage/.

[26] Niklas Masuhr, "AI in Military Enabling Applications," ed. Fabien Merz, *CSS Analyses in Security Policy*, no. 251 (October 2019): 1–4.

[27] "Artificial Intelligence and Life in 2030" (Stanford University, September 2016), https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai100report10032016fnl_singles.pdf.

[28] Zachary Davis, "Artificial Intelligence on the Battlefield – Implications for Deterrence and Surprise," *PRISM: The Journal of Complex Operations* 8, no. 2 (2019).

Therefore, in this case arms control should be interpreted rather loosely –namely as the aim to regulate, manage, or at least monitor developments in this field and, in that sense, to exercise some form of control. Current arms control arrangements still serve as a valuable point of reference, as they reveal various motivations and underlying purposes which may still come into play when discussing RAS. Departing from existing arms control regimes governing weapons that have a lethal, or at least damaging, impact, the first lesson to be drawn is that when it comes to RAS, the context of defensive or offensive use of force is most likely to determine the regulation debate.

Arms control essentially concerns efforts to regulate or limit types and/or numbers of weapons and the ways in which they are used in order to preserve, enhance, or restore international peace and security. Arms control, in conjunction with 'disarmament' and 'non-proliferation'[29] – often referred to as the 'ADN architecture' – has traditionally served strategic interests. Over the last decades, however, humanitarian considerations have surfaced as well. Ethical dimensions of RAS have been researched in much depth, specifically with regard to human agency, human dignity, and responsibility – subjects that relate to such humanitarian concerns.[30] Otherwise, the purposes of arms control have mostly centered around stability (including through disarmament), balance of power, the avoidance of arms races or proliferation beyond state actors, guarantees of a competitive advantage, or the quest for more transparency and predictability through verification.[31] Arms control, whether bilateral or multilateral, ought to be primarily regarded as a risk reduction tool, as well as a confidence and security building measure. One should keep these general notions in mind when answering questions pertaining to the regulation of RAS.

4.3 Accounting for the private sector

Traditionally, much technological innovation has emanated from the military-industrial complex. Later on, these innovations would find civilian applications (known as the *spin-off* effect). In the case of (L)AWS, the trend appears to be going in the reverse direction (*spin-in*), and it is claimed civilian innovation facilitates the design of new weapon systems.[32]

Today, a majority of stakeholders and experts seem to be of the opinion that the civilian domain is 'in the lead' and, therefore, that governments – especially Ministries of Defense (MODs) – are following rather than setting standards. Some experts, though, hold that the application of these innovations for military purposes constitutes a specific next step for which MODs need staffing and capabilities (and to provide context and domain knowledge). In this respect, it may be argued that

[29] Look for instance at NATO definitions: North Atlantic Treaty Organization, "Arms Control, Disarmament and Non-Proliferation in NATO," North Atlantic Treaty Organization, November 28, 2019, http://www.nato.int/cps/en/natohq/topics_48895.htm.

[30] See for example another paper in the framework of this project: Esther Chavannes and Amit Arkhipov-Goyal, "Towards Responsible Autonomy," The Ethics of Robotic and Autonomous Systems in a Military Context (The Hague: The Hague Centre for Strategic Studies, September 2019), https://hcss.nl/sites/default/files/files/reports/Towards%20Responsible%20Autonomy%20-%20The%20Ethics%20of%20RAS%20in%20a%20Military%20Context.pdf.

[31] See for instance John D. Maurer, "The Purposes of Arms Control," *Texas National Security Review* 2, no. 1 (November 2018).

[32] Maaike Verbruggen, "The Role of Civilian Innovation in the Development of Lethal Autonomous Weapon Systems," *Global Policy* 10, no. 3 (September 2019): 338–42.

MODs should be 'launching customers' and engaging in co-development in order to avoid single-vendor dependencies. There seems to be a general consensus that in the framework of designing control regimes, 'spin-in' requires strong interaction with the private sector and will lead to forms of 'shared responsibility and accountability', which are not entirely new but will be more difficult to manage.[33]

Finally, a distinction should be made between civilian innovation as such, and military applications thereof. Governments may no longer be 'in control' of relevant innovation, but they should retain an important developmental role and exercise their convening powers, also with an eye to responsible future control mechanisms.

4.4 Life cycle approach

(L)AWS can be regulated at different stages of the life cycle, from design, production, acquisition, and deployment/use.[34] Considering the particular characteristics of the development of (L) AWS – where the private sector plays a crucial role – a focus on hard law would deal with the consequences, rather than addressing the causes. After all, hard law addresses the final stage of (L) AWS deployment, but leaves unemployed opportunities to engage with key developers at an earlier stage in the life cycle that may help to prevent the 'wrong' use of (L)AWS.

As the private sector rather than the public sector is leading in the design and development of increasingly more dual-use technologies, future regulation must consider the entire system life cycle of research, production, proliferation, development, and use. There is a growing need for 'ethical AI' and industry standards, and at the same time there is significant potential to leverage the private sector and innovation experts for solutions.

4.5 From a rules-based approach to a principles-based approach?

Though not everything about arms control and (L)AWS is new, there is considerable newness to (Lethal) Autonomous Weapons Systems and, therefore, to attempts at regulation. Under the current circumstances, there will be a continuous need for deeper understanding and gradual refinements of regulatory instruments, with the prospect of a larger comprehensive arrangement in the near future being extremely dim. This is due to both the progressive nature of technological development led by the private sector and a diminishing appetite for binding multilateralism among states.

[33] During the aforementioned expert session some doubted in this context the private sector's willingness to engage intensively with governments. The issues of private sector engagement are discussed in more detail in paragraph 5.4 on trusted communities.

[34] For more details, see Esther Chavannes, Klaudia Klonowska, and Tim Sweijs, "Governing Autonomous Weapon Systems: Expanding the Solution Space, from Scoping to Applying" (The Hague: The Hague Centre for Strategic Studies, February 2020).

Returning to the initial question on ‘the need for new norms and arms control’, it seems the answer is affirmative. With evolving values (principles) concerning armed conflict, the norms which derive from these values have evolved too. This will, in turn, be reflected in rules, regulation, and guidelines stemming from these norms. The legal principles of distinction, proportionality, and precaution, as well as norms and rules laid down in control arrangements, will continuously have to be adapted to new types of weapons used in situations of armed conflict. At some point, possibly in the near future, the debate on human-machine interaction may very well lead to the formulation of new principles pertaining to human control, and, concomitantly, to new norms being incorporated into regulatory regimes, as happened after the emergence of weapons of mass destruction. In that sense, taking into account the new quality of (future) (L)AWS and the sequence of formulating values/principles and norms/rules respectively, there is a strong case in favor of approaching the debate primarily on the basis of principles, from which rules may be derived at later stages.

5. Potential paths forward

5.1 The case for hard, soft, *and* voluntary instruments

Building on the insights gained at the expert session held at the Clingendael Institute on 13 November 2019, this section presents a Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis of the three potential levels of regulation introduced above: rules-based 'hard law' instruments, principles-based quasi-legal 'soft' arrangements, and voluntary instruments.

Hard Law. On the positive side, hard law often involves parliamentary ratification procedures (inviting broader public debates and better understanding of the relevant issues) and provides clarity to all parties involved. (L)AWS regulation would become part of a binding body of international law.

However, in the current political climate and considering the fact that the private sector is in the lead with the development of technologies, the negatives seem to outweigh these positives in terms of feasibility. After all, legislative procedures are cumbersome, time-consuming, and will not keep pace with the speed of technological development. Apart from the fact that international law is not always enforceable, the fact that no clear definition of (L)AWS is agreed upon between states, combined with the current pressure on multilateral arrangements in general, makes it unlikely that real progress can be made in rules-based, hard law arrangements (which would anyhow not bind the pertinent category of non-state actors).

Soft and voluntary instruments. Soft and/or voluntary instruments appear to be the more realistic way forward, as these are easier to reach, have lower thresholds for entry, and enable the inclusion of non-governmental stakeholders. Soft arrangements are less static and, by definition, more flexible and adaptable to new circumstances. This approach fits in with a broader development from 'rules-based' to 'principles-based' policies that can also be seen in other fields, such as export control and cybersecurity. Under existing circumstances, this type of arrangement is probably the highest attainable goal.

This is not to disavow the clear downsides to this approach. Some experts rightfully point out that 'talk is cheap' and that the level of adherence will differ in comparison to binding arrangements. Furthermore, changes in political leadership may lead to less sustainable commitments to voluntary agreements, and the appetite for transparency or information exchanges may gradually diminish. With current levels of international distrust, this may dash hopes for satisfactory outcomes, even in this less demanding sphere.

	Strengths	Weaknesses	Opportunities	Threats
Hard	• Binding • Parliamentary involvement • Broad dialogue • Applicable International Law	• Non-enforcement • Time-consuming • (incl. definitions debate) • Participation by and large limited to states	• Incremental development RAS • Civil liability	• Hostile geopolitical environment • No 'ban' foreseen • Potential to hamper civil applications of RAS
Soft	• Easier to agree on: lower barrier • Trust	• Enforcement • Wider participation	• Include non-state actors • Soft law can become wider norms	• Misuse
Voluntary	• Acceptance: low threshold • Wide membership • Builds trust / confidence[35] • De-escalation • Transparency • Self-control • Awareness • Educating the market	• Easy to ignore/quit • Change in political leadership • 'talk is cheap'	• Flexible • Adaptable • Political pressure • Agenda-setting, framing • Basis for further regulation • Trusted data-sharing	• Proliferation of voluntary agreements • Departures when instrument is deemed ineffective

Table 1. SWOT analysis of approaches to regulating (L)AWS.[36]

5.2 Stakeholder involvement

Given the nature of autonomous weapons systems, the involvement of industry merits further attention. By deepening their engagement with the private sector, governments will be able to keep up with technological developments and private sector concerns, empowering them in negotiations on regulation in formal, international settings. At the same time, such engagement can serve as a tool of 'preventive diplomacy' whereby governments can sensitize enterprises operating in the field to a variety of (evolving) concerns with RAS. If key countries and organizations – including the US, NATO and the EU – can lead on ethics domestically, this will set the ground for international principles on military AI and AI in (L)AWS. This is important, as whoever leads on (L)AWS regulations will shape the standards set. For their part, private sector stakeholders will benefit from

[35] This includes the fact that voluntary instruments respond to calls from academic communities for a 'social contract' wherein the MOD gives guarantees with regard to the (peaceful) use of end products developed by those academic communities. On the downside, this potentially puts brakes on the further development by the MOD itself of academic products covered by social contract.

[36] Authors' compilation based on discussions conducted during the expert session.

more appropriate regulation. At the same time, they are encouraged to perform due diligence and self-regulation, as regulation could turn to the positive approach of 'white-listing'.

The key added value of including non-governmental organizations (NGOs) in the debate lies in their role in agenda-setting, expressing public concerns, and, in turn, involving a wider audience in the debate, thereby adding to transparency and, ultimately, more legitimacy to policies. Moreover, serious inclusion of NGOs may also serve to share policy dilemmas and broaden their sometimes limited focus on specific subsets of autonomous weapons – as illustrated by the debate on 'killer robots'.[37] The aim should be to engage all stakeholders and steer the debate away from a crude choice between 'banning' or 'not banning' systems.

The Netherlands might opt to establish new initiatives to exchange information and best practices among key stakeholders, for example through so-called 'trusted communities'. Such voluntary instruments have no formal status but are valuable for their normative and political impact; their role in facilitating information exchange; and ultimately, thereby, their potential to enhance transparency, trust, awareness and accountability. These communities could lead on ethical AI through technical solutions, for example the Ethical Governor, explainable and trustable AI.[38] Although less than ideal in the eyes of many policymakers, who generally prefer more formal instruments, a refocus from a rules-based to a principles-based approach by way of voluntary instruments may be a promising way to address the uncertainties of a system in flux.

Since the idea of 'trusted communities' remains underdeveloped in the Netherlands (and beyond), the next section will discuss them in greater detail. This includes a comparison to more well-known 'epistemic communities'; a discussion of their added value (and pitfalls) compared to other regulatory mechanisms; and the experiences of some countries with similar multi-stakeholder groupings in other emerging tech fields.

5.3 Trusted communities

For several years already, transnational governance literature has pointed to the influential role of non-state actors in the policymaking process.[39] Particularly at times of uncertainty, governments and politicians tend to ask for new and innovative ideas.

[37] Important also because NGO campaigns seem to have a significant impact on public opinion (according to a recent NGO-commissioned YouGov poll, 7 out of 10 Europeans would be in favor of banning 'killer robots': "New European Poll Shows Public Favour Banning Killer Robots," The Campaign To Stop Killer Robots, November 13, 2019, https://www.stopkillerrobots.org/2019/11/new-european-poll-shows-73-favour-banning-killer-robots/.)

[38] Proposed by Ronald Arkin*, the Ethical Governor* is a component of an autonomous robotic system architecture that would prohibit a system from executing an illegal or unethical act prior to it occurring by conducting an evaluation of the ethical appropriateness of any lethal response that has been produced by the robot architecture.

[39] These paragraphs draw on Brigitte Dekker and Maaike Okano-Heijmans, "Emerging Technologies and Competition in the Fourth Industrial Revolution: The Need for New Approaches to Export Control," *Strategic Trade Review* 6, no. 9 (Winter/Spring 2020): 53–67; For transnational governance literature that highlights the role of non-state actors see Mai'a K. Davis Cross, "Re-Thinking Epistemic Communities Twenty Years Later," *Review of International Studies* 38, no. 1 (January 2013): 137–60.

Epistemic communities are the most well-known subset of this, defined as a grouping of scientists linked by their professional ties and ideas in their specific area of expertise.[40] The added value of epistemic communities – even with a small membership – lies in their strong internal cohesion. The extent to which such communities interact with government officials fluctuates, and hence, their influence on the policymaking process also varies – between groups and over time. Other examples of transnational governance groups include transnational advocacy networks and communities of practice.[41] These communities differ from epistemic communities, which are bound together by their knowledge, while transnational advocacy networks are united in their ideals, and communities of practice by their wish to share information.

The narrow definition of epistemic communities has been subject to substantial criticism. In particular, the inclusion of scientists solely in one specific field seems to hamper constructive multidisciplinary solutions, recognition of new trends, and successful translation of knowledge into power.[42] Therefore, the execution of epistemic communities could be extended beyond the current narrow definition. The inclusion of a multidisciplinary team, consisting of businesses, lawyers, government officials and researchers, could lead to discussions among a wide range of expertise and result in a widely shared consensus. The inclusion of government officials would prevent governments from becoming merely rule-takers, and statements deriving from the trusted community could be perceived more legitimately as they would be based on consensus among experts across the field.[43]

Clearly, the creation of 'trusted communities' is a step mainly toward the privatization of transnational governance, stemming from the growing need – and willingness – of both sides for engagement between government and private sector representatives. State actors have less access to necessary technological know-how, complicating any effort to regulate increasingly global challenges, while businesses are more inclined to abide by self-imposed rules of standards, voluntarily setting a precedent for other companies. Motorola Corporation, for example, has effectively contributed to setting telecommunications standards through its chairmanship of the International Telecommunication Union.[44] While critics argue that this trend might transform states from rule-makers into mere rule-takers, and put the level playing field among states at risk,[45]

[40] Peter M. Haas, "Introduction: Epistemic Communities and International Policy Coordination," *International Organizations* 46, no. 1 (Winter 1992): 1–35.

[41] Margaret E. Keck and Kathryn Sikkink, "Transnational Advocacy Networks in International and Regional Politics," *International Social Science Journal* 51, no. 159 (March 1999): 89–101.; and Emanuel Adler and Vincent Pouliot, "International Practices," *International Theory* 3, no. 1 (February 18, 2011): 1–36.

[42] Davis Cross, "Re-Thinking Epistemic Communities Twenty Years Later."

[43] Eleni Tsingou, "Transnational Policy Communities and Financial Governance: The Role of Private Actors in Derivatives Regulation," Working paper (Coventry: Centre for the Study of Globalisation and Regionalisation, January 2003).

[44] John Braithwaite and Peter Drahos, *Global Business Regulation* (Cambridge: Cambridge University Press, 2000), 4.

[45] Peter Utting, "Codes in Context: TNC Regulation in an Era of Dialogues and Partnerships," Briefing (The Corner House, February 2002).

others point out that only private sector firms will have the capacity for research, technology, and development to address and tackle global challenges in the 21st century.[46]

Consultative trusted communities can present an opportunity also for relevant small- and medium-sized enterprises (SMEs) that are often unaware of (possible) uses of their technologies by certain end users. A regular dialogue between representatives of SMEs, start-ups, multinational companies, government officials, and academia active in the field can contribute to information- and best practices-sharing between them.

The success of a trusted community requires substantial and long-term effort. After all, a trusted community depends on a high level of trust between all the actors as dependencies in strategic value chains are increasingly more often exploited. A downside to this path is therefore the significant time and effort involved, especially where communities need to be built from scratch. Separately, care should be taken to avoid a patchwork of parallel trusted communities that complicates business relations, as relevant businesses operating in one sector will also be competitors. Also, trusted communities will inevitably also exclude – and thereby disadvantage – countries or companies 'outside' a trusted community.

Through regular meetings between a fixed membership group, trusted communities contribute to (sensitive) information sharing and best practice exchange in an informal, closed environment. Ultimately, trusted communities serve as a confidence-building and knowledge-sharing instrument that benefits all stakeholders, enhancing understanding and cooperation between government and businesses, as well as discussions on technological developments and (future) regulation. When crafting a new multilateral regime governing (L)AWS, such collaborative fora of organizations will be valuable for sharing lessons learned, preferences, and sensitive information internationally.

5.4 Models to consider: The United States

Although more uncommon in Europe, consultative bodies that bring together a diversity of stakeholders, including businesses, are not a new phenomenon. In Japan, for example, deliberation councils (*shingikai*) have long served as lines for communication between groups – mainly government officials, business representatives, and experts – that operate in distinct but intertwined environments.

[46] Sandrine Tesner and Georg Kell, *The United Nations and Business: A Partnership Recovered* (New York: St. Martin's Press, 2000)., quoted in Peter Utting, "UN-Business Partnerships: Whose Agenda Counts?" (Partnerships for Development or Privatization of the Multilateral System?, Oslo, Norway: United Nations Research Institute for Social Development, 2000), 1–18., (abridged version published in Peter Utting, "UN-Business Partnerships: Whose Agenda Counts?," *The United Nations Research Institute for Social Development Bulletin*, Autumn/Winter 2000.)

The US government has been a front-runner on such trusted communities, having initiated so-called 'communities of caution' that aim to share information on tech-transfer threats.[47] The inclusion of both state and non-state actors in one consultative trusted community has so far, however, been controversial in Europe. A close relationship between the government and industry is historically related to increased industry influence in politics, a practice that long fueled resistance in most European countries (with France as the most obvious exception). While the strict division of business and politics has long proven successful, with the new geopolitical tensions and the rise of emerging technologies, government and industries are experiencing – albeit to various degrees – similar challenges globally. The establishment of trusted communities as a consultative organ consisting of government officials, business representatives, and academia could thus be an answer to the increased overlap between the domains. The inclusion of businesses in the high-level expert group on AI illustrates a new level of openness of the EU and its member states in this regard,[48] but more needs to be done.

Two other examples of trusted communities created by the US government to address similar challenges related to emerging technologies in other fields can be particularly insightful for further Dutch thinking in this field. First among these is the Transglobal Secure Collaboration Participation (TSCP), established in 2002. Initiated by the United States and the United Kingdom, TSCP is a collaborative forum of organizations in the defense industry that enables secure access to sensitive data by creating a cooperative environment based on trust mechanisms. TSCP members comprise government departments and agencies and their prime contractors as well as suppliers, including system integrators and defense manufacturers. The Netherlands' Ministry of Defence is a member of this network. While the focus initially was on secure data access, the TSCP expanded to include data-centric information protection, particularly as a defense against cyber threats.

A second chain of trust that was formed to address challenges stemming from technological development (particularly on export control) is the Emerging Technology Technical Advisory Committee (ETTAC), formed by the US Commerce Department's Bureau of Industry and Security. This kind of strong partnership between government, industry, and academia is particularly valuable now, as an export control regime for emerging and foundational technologies is being established in the United States through the Export Control and Reform Act.[49] The committee's challenge is to create a new regime that "produces the intended benefit of protecting US national security and

[47] Christopher Ashley Ford, "Coalitions of Caution: Building a Global Coalition Against Chinese Technology-Transfer Threats" (FBI-Department of Commerce Conference on Counter-Intelligence and Export Control, Indianapolis, Indiana, September 13, 2018).

[48] "High-Level Expert Group on Artificial Intelligence," European Commission, October 4, 2019, https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence.

[49] Brigitte Dekker and Maaike Okano-Heijmans, "The US–China Trade–Tech Stand-Off" (The Hague: The Clingendael Institute, August 2019).

promoting US technical leadership without compromising US economic competitiveness or even unwittingly undermining that same technical leadership."[50]

The Dutch government could apply this model by identifying the key domestic stakeholders in the domain of (L)AWS and facilitate the creation of stronger networks through these multidimensional forums. Through trusted communities, industry and academia can provide input for suitable adjustments to the RAS regimes and can cooperatively balance innovation, economic benefits, and security. At the same time, the Dutch government can share with these stakeholders new and evolving concerns with regard to the development, review, and deployment of such systems. This can help raise awareness of international political dynamics among high-tech start-ups and small- and medium-sized enterprises that may be unaware of the potential (mis)use by certain players of their technologies.

To some extent, the Dutch government is already facilitating such trusted communities in related fields, such as export control, and digitalization and ethics,[51] while consultation rounds organized prior to international meetings on cyber security also fit in with this trend. With regard to RAS, encouraged by a motion adopted in the Dutch Parliament, the Dutch government has started to reach out to the private sector – as exemplified by a speech by Foreign Minister Stef Blok to drone manufacturers at Amsterdam Drone Week. The minister invited them "to brainstorm with me about solutions to an urgent and complex issue. I want you to use those solutions to change the world," adding that "My aim is for our joint efforts to foster a global alliance of international policymakers and companies [of drone-producing countries] that commit themselves to sharing ideas and developing practical standards ensuring that commercial drones are used peacefully. This alliance would enable us to maximize the potential of drone technology as a force for good."[52]

Setting up a trusted community is one thing; deciding what topics to put on the agenda is another. Discussions should not merely copy those within the CCW/GGE – with the only difference being that states would be less, and industry and other stakeholders would be more prominently represented – but can certainly elaborate on issues identified in that framework. Take, for instance, the GGE's latest guiding principle on human-machine interaction (see page 9). It is easy to discern the various elements contained in this principle and to establish whose input must be ensured: the 'various stages of the life cycle of a weapon' requires industry's judgments, while making certain that the use of weaponry is 'in compliance with international law' is the remit of governments. Similarly, concerning the 'quality and extent of human-machine interaction', the 'operational

[50] Stephen Ezell and Caleb Foote, "How Stringent Export Controls on Emerging Technologies Would Harm the U.S. Economy" (Washington, DC: Information Technology & Innovation Foundation, May 2019).

[51] Two such examples are the emerging tech export control expert session that was held on 16 December at the Clingendael Institute, The Hague; and "Aanpak Begeleidingsethiek in Het NRC," ECP: Platform voor de InformatieSamenleving, December 5, 2019, https://ecp.nl/actueel/aanpak-begeleidingsethiek-in-het-nrc/.

[52] Stef Blok, "Speech by Stef Blok, Minister of Foreign Affairs, at Amsterdam Drone Week, 6 December 2019" (Amsterdam Drone Week, Amsterdam, December 6, 2019), https://www.government.nl/documents/speeches/2019/12/06/speech-by-minister-stef-blok-at-amsterdam-drone-week.

context' is to be provided by the armed forces, while the 'characteristics and capabilities' of the weapons under discussion would again allude to manufacturers. This is but one example of how, in a trusted environment, some of these issues can be elevated and, subsequently, inform the overall debate.

Finally, trusted communities should facilitate long-term goals and adopt structural characteristics as opposed to being crafted as high-level meetings held on an ad hoc or one-off basis. Autonomous weapons regulation is by definition a longer-term issue that will transcend the ebbs and flows of government terms and require ever deeper understanding. In that sense, this topic seems to be eligible for further elaboration in trusted communities.

6. Conclusion

In recent years, geopolitical tensions and rapid technological developments have increased, and the international system of arms control and international trade of military items has been under pressure. This should concern the Netherlands as a country traditionally supportive of the rule of international law and multilateral agreement in general. It should also be of concern because of a growing domestic debate, both inside and outside parliament, on the use of autonomous weapons systems in armed conflict.

As hard law arrangements become more difficult to negotiate and to uphold, and regulators are increasingly less able to keep up with the rapid technological developments, (L)AWS regulation requires new approaches and new instruments. Apart from the continuation of efforts in formal frameworks (such as the CCW or the Wassenaar Arrangement on the export of dual-use items), the government may also reach out to businesses, knowledge institutes, lawyers, and other stakeholders at home as well as in like-minded states through trusted communities that can be helpful in enhancing the debate. Such networks have the ability to bring together key actors to provide input for developing principles and norms for further regulation or export control regimes that are based on mutual trust and respect.

The Netherlands government has already embarked on a similar road with the announcement of an international conference in 2020 that involves partner countries, industry experts, and NGOs on the responsible development and use of armed unmanned aerial vehicles.[53] Depending on the outcomes of this initiative, such activities may also be (co-)organized on topics more generally related to the overall debate on autonomous weapon systems. It would be fitting for the Netherlands, traditionally a champion of multilateralism and arms control, to remain actively engaged with this matter and, be a representative of the 'middle ground' within the CCW that advocates for a balance between a ban and unfettered proliferation. The Dutch are in a unique position to utilize available experience and knowledge to shape further discussion.

[53] Stef Blok, "Betreft Motie Koopmans c.s. over Beheersing van de Productie, Plaatsing, Verspreiding En Inzet van Nieuwe Potentiële Massavernietigingswapens" (Ministerie van Buitenlandse Zaken, September 20, 2019), https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2019/09/20/kamerbrief-over-nieuwe-potentiele-massavernietigingswapens/kamerbrief-over-nieuwe-potentiele-massavernietigingswapens.pdf.

Based on the above analysis, some route markers may be listed pertaining to the initial questions about the need for regulation of RAS, the forms which this may take, and the ways in which to organize such efforts:

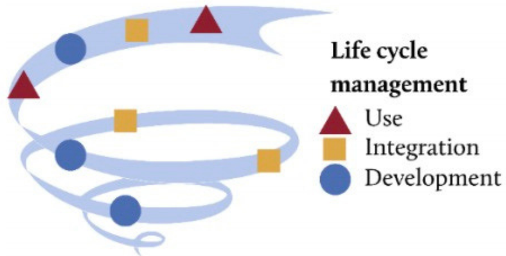
- The relevant category of autonomous weapons to be addressed is located in the cross section between high- and full-level automation and the defensive and/or offensive use-of-force functions, the *effects* of which are to be regulated*;*
- The considerations motivating control of autonomous weapons are not likely to differ much from the purposes behind existing arms control arrangements (such as risk reduction, confidence and security building, preserving international peace and security out of both strategic and humanitarian concerns);
- Given the nature of these new technologies, regulation should comprise the design phase of systems and incorporate the context for their incremental introduction and use by the military. This presupposes early and comprehensive collaboration between industry and end users, while the latter should seek to retain a leading role;
- The answer to the question whether additional regulation is required seems to be affirmative as there is a continuous need for deeper understanding and gradual refinements of existing regulatory instruments, while the prospect of a comprehensive arrangement in the near future remains dim;
- Parties involved should be aware of the fact that the debate on autonomous weapons is part of a larger debate on technology and human-machine interaction that only recently has been gaining traction;
- The ongoing formal debate within the CCW/GGE is necessary, but probably not sufficient to advance regulation. This suggests that alternative fora are needed to move from ‘thin state consent’ to ‘thick stakeholder consensus’;
- At this stage, principles-based discussions are required next to rules-based debates, while there seems to be an expressed preference to focus on softer and/or voluntary instruments (as opposed to legally binding agreements);
- Because of the potential downsides and the unlikelihood of prohibitive measures, a more promising strategy may be to work on white-listing (drawing up *dos* instead of *don’ts*);
- The mostly civilian sources of technological innovation and emerging ‘spin-in’ effects, as well as shared levels of responsibility (and/or accountability), would also argue in favor of a multi-stakeholder approach for deliberation;
- In this regard, the development of ‘trusted communities’ as tools for transnational governance could offer a promising way forward, and the Netherlands may draw lessons from the experiences of other countries in related emerging tech fields.

Factsheet

Effective Stakeholder Cooperation during the Lifecycle of Robotic and Autonomous Systems

Stakeholder Cooperation

The emergence of RAS poses new questions and challenges to the effectiveness of cooperation with various stakeholders during the life cycle of RAS.



Challenges

There are distinct elements that make the development, integration and use of RAS - and the interactions/ cooperation that comes with them - more challenging:

RAS development is mainly driven by civilian innovation, instead of by traditional defense industry

RAS must be developed in fast procedures, used for shorter periods and updated throughout it's life cycle

Ethical and legal debate around RAS requires interaction with a wide range of external stakeholders

The fundamental changes that go hand in hand with RAS require broad interaction with a range of stakeholders

Recommendations

Against this backdrop, the following recommendations are made for the armed forces in order to improve cooperation with a range of stakeholders regarding RAS:

Development

1. Establish close communication in the early-design phase of development
2. Ensure division of tasks and responsibilities with external stakeholders
3. Create incentives for parties to enter the RAS scene while fostering trust among partners

Integration

1. Develop training programs in cooperation with RAS developers
2. Test RAS in an as-close-as-possible-to-reality operational environment
3. Accompany RAS handover with division of future tasks, responsibilities & accountability
4. Incorporate internal stakeholders early in the integration process

Use

1. Outline clear division of responsibilities and control
2. Conduct post-mission evaluations and reflect on successes and failures
3. Further integrate third party private actors into the upgrading of systems
4. Provide personnel and partners with training on cyber awareness
5. Consider an incentive based scheme to mitigate unintended risks

The Hague Centre for Strategic Studies

The Hague Centre
for Strategic Studies
Lange Voorhout 1
2514 EA The Hague
The Netherlands
hcss.nl

Capstone Report
Robotic and Autonomous Systems in a Military Context

The Hague Centre for Strategic Studies

Chapter 4

Effective Stakeholder Cooperation during the Lifecycle of Robotic and Autonomous Systems

Bianca Torossian, Frank Bekkers & Klaudia Klonowska

1. Introduction

It is commonly accepted that armed forces cooperate with external actors, outsource the production of arms, and share Research & Development projects with private companies and universities. Even though this practice is long-standing, the emergence of Robotic and Autonomous Systems (RAS) poses new questions and challenges to the effectiveness of multi-stakeholder cooperation in a military context. RAS are unique in that they ultimately can take humans 'out of the loop'[1] and, as a consequence, drastically affect operational performance, organizational embedding (e.g. influencing numbers, skills and training of personnel), operational concepts (i.e. doctrine and tactics), and raise specific ethical and regulatory concerns. In short, the introduction of a new RAS in the armed forces is seldom a 1-1 replacement of a more human-centric solution or a seamless fit to an identified capability gap. The nature of RAS intervention is disruptive and therefore renders the interaction between stakeholders more complex.

This paper studies the emerging complexity of relations between a wide variety of stakeholders involved in the development, integration and use of military RAS. The focus lies specifically with the interactions between the military and private parties, i.e. industry, knowledge institutes, and civil society (although the role of the national and international policy makers in this process is also acknowledged). Based on the findings, best practices are outlined and key requirements to improve the effectiveness of cooperation are highlighted through organizational, legal, and practical solutions.[2]

[1] The scale from human 'in the loop, 'on the loop', to 'out of the loop' refers to the degree to which a human is involved in the operating and/or decision-making process of the system.

[2] The analysis provided in this paper is the result of an analysis of relevant literature, strengthened by insights from the expert session held by The Hague Centre for Strategic Studies on the 13th of February 2020. This session gathered an interdisciplinary group of professionals representing the government, businesses, and knowledge-institutes. Insights from the session are integrated throughout the paper.

This paper is organized in the order of basic system life cycle stages: (1) development, (2) integration (transfer of ownership, organizational embedding), and (3) use of RAS in an operational environment. However, the specific nature of RAS often results in a Concept Development & Experimentation (CD&E) process, in which successive phases of development, acquisition, initial introduction, and use form a spiral development process (see Figure 1).

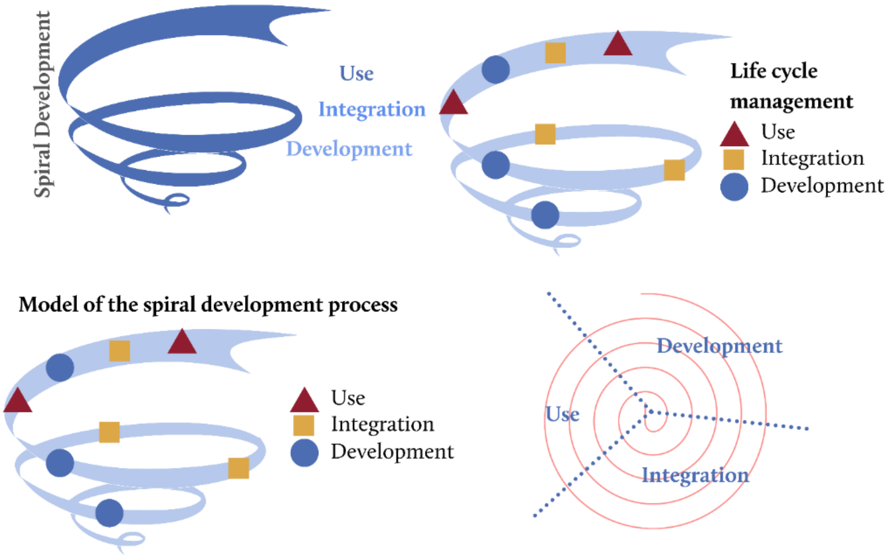


Figure 1. Spiral development process and the reoccurring stages of the system life cycle.

During the development and integration phases, concept development and testing of successive prototype versions go hand-in-hand to mature the system, potentially including regular revisions of system requirements. During the use phases, lessons are learned that may retrospectively influence the functionality and design of a system, thereby requiring (potentially radical) updates. This iterative process blurs the distinction between each phase, as well as the lines between the developer/producer of the system on the one hand, and the defense organization as customer and user on the other. Instead of a definite hand-over in terms of ownership, responsibilities and liabilities, the various stakeholders are typically involved and interconnected throughout all stages of the RAS life cycle. Therefore, this paper also discusses the overarching issues and solutions under the heading of 'RAS life cycle management'.

Accordingly, chapter 1 outlines key differences between multi-stakeholder cooperation with RAS and other military technologies; chapters 2 to 4 discuss in further detail requirements of cooperation at each stage of the RAS life cycle, that is, development, integration and use; while chapter 5 provides a discussion of the overarching requirements for the so-called 'life cycle management'. The concluding chapter provides recommendations for effective and continuous cooperation between various stakeholders working with RAS in the military context.

2. Distinct cooperation challenges for RAS

There are a number of elements that make the development, introduction and use of RAS—and the stakeholder interactions that come with it—different from more traditional (linear) capability development processes:

- The development of RAS is to a great extent driven by civilian innovation, thus creating demands for interaction with designers, developers, and manufacturers outside the traditional defense industry;[3]
- Due to a rapid cycle of innovation within e.g. artificial intelligence (AI), RAS must be developed and acquired in fast-paced procedures, used for shorter periods of time, and modified, updated, inserted, or exchanged throughout the life cycle;
- The ethical questions and legal uncertainties surrounding the use of unmanned and increasingly autonomous systems demands interaction with a range of stakeholders and policy makers external to defense organizations;
- The fundamental changes that RAS might bring to (some or all of) the DOTMLPF-elements[4] requires broad interaction with stakeholders within the defense organization, with international military partners, and possibly with other partner agencies.

These distinct characteristics, risks, and opportunities associated with RAS give rise to specific requirements for cooperation between the armed forces and non-military partners to ensure effective development, integration and use of RAS.

[3] Traditionally, technological innovation has emanated from the military-industrial complex. These innovations would later find civilian applications (known as the spin-off effect). In the case of RAS, as for other military systems that derive a large part of their functionality from information technology, the trend goes in the reverse direction (spin-in). Verbruggen, "The Role of Civilian Innovation in the Development of Lethal Autonomous Weapon Systems," 338–42.

[4] Doctrine, Organization, Training, Materiel, Leadership, Personnel and Facilities. See, for example, https://military.wikia.org/wiki/DOTMLPF.

3. Development

The development of RAS is a dynamic process of hardware and software design and production, which at later stages is consistently revisited according to the results of system testing, integration, monitoring, and use. It is increasingly more common for military forces to outsource the development of technologies to private actors in order to acquire unique expertise and skills needed for the development of sophisticated equipment such as RAS.[5] Innovations and a rapid growth in the public sector in the use of unmanned and increasingly autonomous systems (e.g. medical robots, autonomous vehicles, and an abundance of civil drones) inspire the adaptation of civil platforms for specific military purposes.[6] Thus, it can be said that RAS in the military context is developed in a 'spin-in' environment, whereby the civil domain leads in innovation. This chapter outlines solutions to improve the effectiveness of cooperation in this dynamic environment during the development of RAS.

3.1 Integrated and interdisciplinary cooperation

The first step in envisioning an effective cooperation throughout the lifecycle is to establish a suitable stakeholder cooperation at the initial phase of RAS development. RAS development demands the involvement of an increasingly interdisciplinary and long-term approach to stakeholder cooperation. Studies show that numerous states have already conducted interdisciplinary RAS-related projects in cooperation with industrial consortiums, universities, laboratories, and start-ups.[7] Interdisciplinary teams promise to meet the demand to adequately converge 'technical' and 'conceptual' requirements due to the integration of knowledge of actors with military experience on the one hand, and technical skills on the other. Additionally, it is desirable for knowledge institutes to be involved at this stage to provide out-of-the-box insights and for civil society to highlight possible ethical concerns. Cooperation at this stage should be extended and, where possible, well-integrated throughout the entire life cycle.

[5] Slijper, Beck, and Kayser, "State of AI: Artificial Intelligence, the Military and Increasingly Autonomous Weapons."
[6] Perlo-Freeman and Sköns, "The Private Military Services Industry," 4; Verbruggen, "The Role of Civilian Innovation in the Development of Lethal Autonomous Weapon Systems."
[7] Slijper, Beck, and Kayser, "State of AI: Artificial Intelligence, the Military and Increasingly Autonomous Weapons."

3.2 Division of tasks, investments and responsibilities

Cooperation between and co-creation of various stakeholders requires managing where responsibilities lie. One way to manage the extent of private actor involvement in RAS development is through contracting. The military should carefully consider how to establish a long-term relationship with the RAS developer/producer that will extend into other stages of the RAS life cycle on the one hand, and the ability to change partnership in case of ill-performance on the other. Under the typical co-creation conditions of CD&E, questions like 'who does what', 'who pays for what' and 'who bears responsibly for what' do not always have a clear-cut answer. Contractual arrangements with built-in flexibility may provide guidance.

Cooperation with other countries further exacerbates the above-mentioned need to clearly divide tasks, investments, and responsibilities in RAS development. With the increase of contractors in the development and production of RAS, there are further difficulties in ensuring adequate design, oversight, and quality. Possible challenges include inconsistent communication (e.g. incoherent testing results may result in faulty design), lack of independent reviewers, or misleading political agendas. Effective solutions require political authorities and armed forces to harmonize requirements through, for example, tailored Memoranda of Understanding and international legal instruments.[8]

3.3 System architecture

The effectiveness of RAS depends greatly on the ability to communicate and share information with other sensors in order to collect and process real-time events in a dynamic environment.[9] Additionally, the ability of the military to respond and mitigate risks is dependent on the design of interface. From the technical perspective, this requires that the developer is progressive and can provide the most advanced solutions to create "easier, safer, and more flexible" systems.[10] From the military perspective, the chosen architecture must ensure that RAS can communicate with systems of other units or other countries in order to provide interoperability.[11] Common system architecture entails trusted intelligence sharing when cooperating with friendly forces, and security of information to prevent data exploitation by adversary forces. The choices made regarding the system architecture fundamentally influence the RAS (inter)operability, thus cooperation between all parties is required in order to pose and answer adequate questions.[12]

[8] Clarke, "The Arrow Missile: The United States, Israel and Strategic Cooperation," 478.

[9] Kortenkamp and Simmons, "Robotic Systems Architectures and Programming," 188; "Robotic and Autonomous Systems of Systems Architecture," 38–39.

[10] Kortenkamp and Simmons, "Robotic Systems Architectures and Programming," 202.

[11] "Robotic and Autonomous Systems of Systems Architecture," 37.

[12] For the type of questions that relate to the system architecture both from the technical and organizational perspectives, see Kortenkamp and Simmons, "Robotic Systems Architectures and Programming," 202–3.

Besides interoperability, another consideration for the army is the ability to insert new RAS into the already existing platform-based systems. Changing demands of the army require that systems have modular and flexible architecture that can “facilitate the insertion of independently developed ‘best-of-breed’ cognitive functionality”.[13] In order for the developer to respond to these challenges, it is necessary for them to know in advance what the current state-of-the-art of military systems are and what (future) demands need to be met.

3.4 Matching ‘pull’ and ‘push’ factors

Matching military requirements with technological possibilities in a situation where both are moving targets presents a challenge. Military users and technical developers have different frames of reference and must therefore collaborate closely to ensure proper translation of military demands into technical parameters (and vice versa) in an ongoing spiral development process. Effectively balancing military requirements and expectations, technological possibilities, and (potentially conflicting) legal, ethical and safety parameters, is a requirement that all involved stakeholders must manage in order to facilitate cooperation.

[13] “Robotic and Autonomous Systems of Systems Architecture,” 37.

4. Integration

Integration concerns the organizational embedding of mature RAS, of systems that are selected to be scaled-up from an experimental setting to the structural use. At this stage, the relationship with the developer/producer of the system changes and new actors, such as the actual military end-users, emerge or acquire a more dominant role. During this stage, the nature of 'hand-over' changes and raises new questions regarding the role of each actor. This chapter highlights the changing nature of integration of systems in the military context and proposes ways to enhance multi-stakeholder cooperation.

4.1 'Hand-over' of military systems

The integration stage includes a type of 'hand-over' from the developer/producer to the military. However, a feature of RAS is the dependency upon integrated software that continuously evolves; certainly, where self-learning algorithms are part of the autonomous reasoning of the system. As a result, the hand-over of RAS does not necessarily finalize the involvement of the producer in the latter stages of the life cycle. The responsibility is likely to extend for the producer to ensure that the system is adequately and regularly updated, and that the self-learning nature of the system is controlled and continues to meet demands and standards. If the changes made by self-learning algorithms lead to alterations in the use of a system, this should be well explained to the operator prior to the first use after an 'update'.

Therefore, the hand-over of the system should be accompanied with a clear division of future responsibilities and liabilities, as well as the accountability for system failures, servicing, and software updates. A proposed alternative to the traditional hand-over of ownership is the licensing model or the service model, in which the developer/producer remains the owner of the system. Though licensing agreements may set strict criteria to guarantee the safety of a system, concerns may be raised regarding their exclusive ownership and control by private actors, since conflicts of interest can arise between the efficiency of the systems on one hand and national security or public safety on the other.

4.2 Testing environment

It is evident that all military equipment needs to be tested in order to provide quality assurance. In the case of RAS, it is particularly important to ensure that the system is tested in an environment that reflects, as close as possible, the intended operational environment. In order to meet the demands, it is desirable for a military end-user to have the opportunity to interact with the system and its interface in order to better understand the (autonomous) functionalities, its limitations and possibilities, and to develop a sense of trust in the system. Observations gathered from such an exercise, when conducted in collaboration with the developers, are likely to yield new upgrades in the system.

4.3 Comprehensive integration

The integration of RAS into military forces requires adaptation of processes beyond the units where RAS are deployed. It involves the adaptation of all the 'DOTMLPF' categories.[14] The military should consider whether the doctrine covers situations of RAS deployment, whether the training and organization of forces are sufficient to ensure that RAS are taken full advantage of, whether there is sufficient technical literacy to deal with ad-hoc technical problems, whether the facilities are equipped to repair RAS, etc. These questions should be answered during the integration stage in cooperation with actors that are involved in the development, *as well as*, with the end-users.

[14] Doctrine, Organization, Training, Materiel, Leadership, Personnel and Facilities. See, for example, https://military.wikia.org/wiki/DOTMLPF.

5. Use

The use stage of RAS involves the deployment, maintenance, and service of RAS. The use of RAS in operational environments fundamentally influences the ways in which the military work: how it conducts missions, with whom and under what conditions. Among others, this fundamental change can be attributed to operators and commanders interacting with the system at "higher levels of abstraction".[15] Specific solutions to improve cooperation between stakeholders at this stage are outlined below.

5.1 Human-machine teaming

RAS are typically deployed as part of mixed human-machine teams. Soldiers and RAS work alongside under high-demand circumstances, with mutual trust as a condition *sine qua non*. Personnel working with and alongside RAS in actual operations must be willing to adapt to changing circumstances and improve the understanding of and trust in system functionalities before deployment. Their tasks should be clearly divided to determine which team members are responsible for the validation of targets and have the control to override the decisions of the system. A clear division of responsibilities helps to reduce automation bias, promotes compliance with international legal obligations and principles, ensures human intervention, and prevents mode confusion (a situation when operators erroneously switch between highly- and less-automated modes).[16]

After use of RAS in an operational environment, it is important to reflect upon ways in which the system aided the operation, as well as failures to add value to mission objectives. An integrated process of feedback and improvement aids the culture of 'shared risk', whereby responsibilities, interests, and values are shared among stakeholders. Considerations from this stage of evaluation should be communicated to relevant stakeholders and/or to personnel that interact with similar RAS in other operational environments.

5.2 Continuous technology insertion

Since the development of, for example, AI-related technologies is fast-paced, RAS requires a (relatively) rapid update and upgrade cycle to stay relevant and competitive. Therefore, besides regular service and maintenance of RAS, third parties may continue to be involved in the re-configuration, updating, and

[15] Platts, Cummings, and Kerr, "Applicability of STANAG 4586 to Future Unmanned Aerial Vehicles," 2.
[16] Platts, Cummings, and Kerr, 12.

upgrading of system functions after deployment.[17] Collaboration with stakeholders should be logically integrated in order to ensure that RAS upgrades continue to meet the demands of the military, are clearly understood and trusted by operators and other relevant personnel, and incorporate feedback provided from previous deployments. If necessary, from the upgraded system functionalities, military personnel should receive additional training to ensure their ability to interact with the system.

5.3 Security

RAS face specific operational security challenges, both in terms of intentional cyber/data breaches as well as unintentional failures. Parallel to the development of RAS, there are actors advancing capabilities to intercept, disrupt, jam, spoof, and hack communications of robotic systems.[18] Additionally, RAS relying on data that is stored remotely poses new risks to the management of sensitive data.[19] Even with cryptography providing new ways to encrypt data while maintaining confidentiality and integrity of information,[20] new questions arise of control over sensitive data and related infrastructure. Further efforts are necessary to provide military personnel and relevant partners with adequate cyber awareness trainings specific to the additional risks associated with the use of RAS.[21] Due to the greater level of technical understanding of system amongst the developers, it is desirable for private actors to inform the military about the requirements for effective use of their systems.

At the same time, the use of RAS may lead to unintentional safety issues. The machine learning techniques embedded in RAS may result in unpredictable and harmful accidents, due to an incorrect specification of functional objectives, inconsistent oversight over the learning process or other implementation errors.[22] It has been proven that even the most careful designs of automation known to us today may lead to 'system accidents'.[23] It is desirable for the military to continue cooperating closely with the initial designers of RAS in order to monitor and correct for system redundancy, automation bias, and neglectful responses.[24] An incentive-based scheme (similar to insurance schemes that reward safe behavior) may be used to reward the designers for implementing risk aversion measures.[25] The implementation of these so-called 'fail-safe' measures may help to mitigate unintended risks. Furthermore, it is important to remember that safety of RAS depends on the extent of system unpredictability, which in turn depends upon the extent of allowed automation in military technologies by the political authorities. Therefore, in order to mitigate the risks of RAS, it is necessary to keep an open and informative dialogue with the political decision-makers.

[17] Kortenkamp and Simmons, "Robotic Systems Architectures and Programming."
[18] "The U.S. Army: Robotic and Autonomous Systems Strategy," 15.
[19] Bromley and Maletta, "The Challenge of Software and Technology Transfers to Non-Proliferation Efforts"; Allen and Chan, "Artificial Intelligence and National Security."
[20] Allen and Chan, "Artificial Intelligence and National Security," 91.
[21] "Army Cyber Training and Education within Finabel Member States," 11 (Pre-mission cyber awareness refers to the understanding of cyber threats within a specific new environment as well as other permanent threats such as social media.).
[22] "Safety, Unintentional Risk and Accidents," 3.
[23] "Safety, Unintentional Risk and Accidents," 3.
[24] "Safety, Unintentional Risk and Accidents," 12.
[25] See Wasiak, "What Is the Incentive in Insurance Premiums?"

6. Life cycle management

As previously stated, the development, integration and use of RAS is not linear due to its evolving nature which requires an iteration of military requirements, review of technical parameters and regular adjustments to allow for technological progress and new operational insights to be reflected in the system and its actual use. This spiral development process typically requires long-lasting and multifaceted relationships between the developer/producer and the military user. The dependency that arises consequently calls for effective arrangements between the parties involved that include the entirety of the life cycle. Accordingly, the following requirements are particularly relevant to the life cycle management of RAS.

6.1 Meaningful oversight over stakeholders

The distinct nature of RAS raises regulatory and ethical questions. Studies have indicated a difficulty in determining accountability and liability for the wrongdoings caused by RAS.[26] With a greater number of actors involved, additional measures are necessary to ensure the line of accountability is clearly defined. This is particularly salient when cooperating with private companies, as this interaction increases the 'distance' between the decision-making and the operational behavior (potentially including the use of force). Even subtle differences in the design of a system may have critical consequences on the ways that RAS behave or the military interacts with RAS in an operational environment.[27] The notion of accountability is challenged when private entities are involved since RAS operate based on the parameters and functions set by private actors, and international law is centered around the principle that states alone have "the exclusive legitimacy to exercise violence".[28]

Thus, the military should exercise meaningful oversight over all stakeholders, including private contractors, in order to ensure the implementation of relevant legal restrictions and ethical considerations specified either in international or domestic instruments. The basis of such oversight mechanisms can take a variety of forms, from internal guidelines, to codes of conduct and external harmonized international NATO frameworks, and should always be accompanied by communication mechanisms in order to ensure adequate interpretation and application.

[26] Chavannes, Klonowska, and Sweijs, "Governing Autonomous Weapon Systems."
[27] Chavannes and Arkhipov-Goyal, "Towards Responsible Autonomy: The Ethics of Robotic and Autonomous Systems in a Military Context," 19.
[28] Perlo-Freeman and Sköns, "The Private Military Services Industry," 13.

6.2 Non-proliferation

The difficulty of adequately regulating RAS proliferation is further exacerbated by the large quantity of actors involved in the RAS life cycle. There is a possibility that private companies may re-sell or re-use RAS after the expiration of the contract or that states may transfer RAS elements to third countries without the approval of production partners.[29] There is a need for states to prevent proliferation of RAS and its elements by external partners beyond the established cooperation; for example, to ensure that the period of non-engagement is respected and the success of the project, and ultimately national security, are not endangered.

6.3 Collaborative engagement

An important requirement that may not be easily translated in contractual terms or technical parameters, but is nevertheless critical to an effective cooperation, is the creation of the so-called 'Formula 1 mindset'. The F1 mindset refers to a common sense of urgency shared by a diverse team that orients itself toward a clearly defined goal. A Formula 1 racing team is not only about the driver at the forefront of all the attention, it is about the way in which the whole team collaborates to achieve a shared goal of being the best and the fastest. Every role and member are as important as any other within the team. With a growing number of actors involved along the life cycle of RAS, there is a need to ensure that all parties work toward a common goal that stands above the interests of particular stakeholder affiliations and is integrated under one team's effort. The F1 mindset should include the political authorities, whose support and will are essential to the effectiveness of missions that deploy RAS.

6.4 Managing private partnerships (PPP) and market competition

One of the challenges of the military contractor is to balance the cooperation with reliable and trustworthy partners and the management of supply from new innovative competitors. Long-lasting arrangements may result in high dependency on a single provider, who may use his pivotal position to "increase charges and lower quality" while minimizing the possibility of effective oversight.[30] Balancing the risks and benefits associated with both market competition and close PPP is needed to achieve 'the best of both worlds' during the entire life cycle of RAS. One of the ways in which this situation is being addressed in the United States is through open bidding acquisition processes, accompanied by an additional process of scouting revolutionary and emerging tech companies in order to increase the number of providers.[31] At the same time, due to the complexity and risks related to the development of RAS, it is often a preferred option to work with trusted partners, whose long-standing performance and relationship provides a sense of confidence.

[29] Clarke, "The Arrow Missile: The United States, Israel and Strategic Cooperation," 483 ("Israel has employed US weaponry contrary to US law and policy, incorporated US technology into Israeli weapons systems without prior approval, and made improper transfers of US missile and other defense systems and technologies to other countries, including Chile, China, and South Africa.").

[30] Perlo-Freeman and Sköns, "The Private Military Services Industry," 15.

[31] Allen and Chan, "Artificial Intelligence and National Security," 14 (An example is IN-Q-TEL company that searches for additional companies to participate in the governmental contracts.) .

7. Recommendations

This study reveals that in order to achieve an effective cooperation along the life cycle of RAS long-lasting relationships are required, such that are supported by clearly defined common goals, harmonized procedures, and shared understanding. This study highlights organizational, legal, and practical solutions to improve the effectiveness of cooperation with RAS. It is apparent that the dependency of RAS functionalities on the initial design is critical, thus highlighting the need to integrate and manage the cooperation in an iterative process that includes all parties, from the front developers to the end-users.

Based on the above-outlined considerations, the following recommendations are made to the armed forces in order to improve cooperation with a range of different stakeholders regarding RAS.

Development

1. *Establish* close communication from the early design phase onwards between diverse stakeholders (i.e. military customers and users, engineers, academics, civil society representatives, legal/ ethical experts, and policy-makers) in order to improve implementation of interdisciplinary considerations into the functional and technical parameters of RAS.
2. *Ensure* that the contracting arrangement with external parties includes a clear division of tasks and responsibilities from the early design phase. The military should carefully consider how to establish a long-term, trusted relationships with the RAS developer/producer that will extend into other stages of the RAS life cycle on the one hand, whilst maintaining the ability to change partnership in case of ill-performance on the other.
3. *Create* incentives for new parties to enter the military RAS scene while fostering trust amongst current partners. This could take the form of offering fellowships to academic researchers and private entities to continue innovating RAS of autonomous systems, the organization of national competitions promoting participation of researchers and engineers, or offering financial incentives (such as tax reliefs or state funds) to private companies to encourage cooperation with the public sector in the creation of RAS.

Integration

1. *Develop* effective training programs in cooperation with RAS developers in order for personnel to gain an understanding of the system, to improve technical literacy, to build trust, and to adjust team dynamics.
2. *Test* RAS in an environment that is as close as possible to the intended operational environment in order to provide quality reassurance and improve users' trust in and understanding of the system (i.e. functions, limitations and possibilities).
3. *Accompany* the hand-over of the system with a clear division of future responsibilities, accountability for system failures, servicing, and software updates. This could be realized through the creation of licensing models in which the developer remains the owner and the party responsible for upgrades and quality checks.
4. *Incorporate* a wider community of internal stakeholders early in the process to assure that necessary organizational, procedural, and doctrinal changes that stretch beyond the scope of actual units in which RAS are introduced. For example, RAS may involve the hiring of different sorts of personnel, tactics and procedures that affect a wider operational deployment.

Use

1. *Outline* a clear division of responsibilities and control to help to reduce automation bias, promote compliance with international legal obligations and prevent 'mode confusion'.
2. *Conduct* post-operation evaluations to reflect upon ways in which the system aided the operation, as well as, failures to serve mission objectives. Considerations should be communicated with relevant stakeholders, including developers and designers of RAS, and personnel who interact with similar RAS in other operational environments.
3. *Further integrate* third party private actors into the re-configuration and upgrading of system functions after deployment.
4. *Provide* military personnel and relevant partners with adequate cyber awareness trainings specific to the additional risks associated with the use of RAS.
5. *Consider* an incentive-based scheme, in which higher safety measures and risk aversion are rewarded to mitigate unintended risks.

Life cycle management

1. *Agree* upon a clear division of tasks, investments and responsibilities between stakeholders that among others include, hand-over arrangements, quality control, liability, servicing, maintenance, and software updates.

2. *Exercise* meaningful oversight over all stakeholders, including private contractors, in order to ensure the implementation of relevant legal and ethical restrictions. This may take the form of internal guidelines or codes of conduct, supported by strong communication mechanisms.

3. *Prevent* proliferation of RAS elements by partners beyond the established cooperation through auditing and enforcement of contracting responsibilities.

4. *Promote* a 'Formula 1 mindset' between developers, technicians, politicians, and end-users of RAS whereby a common sense of urgency is shared by a diverse team that orients itself toward a clearly defined goal.

Factsheet

The Implementation of Robotic and Autonomous Systems: The Future is Now, Prepare for 2045

The Future Implementation of RAS

Robotic and Autonomous Systems (RAS) are here to stay. How can the Royal Netherlands Army (RNA) use such systems in the future? And how can RAS best be implemented towards 2035 and 2045?

Looking Into the Future: "The Era of Relentless Competition"

The year is 2045, the Great Power Competition that started in the late 2010's has exacerbated over the last 25 years, resulting in high tensions globally. Dutch Armed Forces are deployed to the Nile delta, where they are tasked with border security and preventing human trafficking, smuggling and data crimes. The challenges on the ground are considerable: the opponent has acquired advanced technologies in unmanned and Artificial Intelligence,relying on the force and fearlessness of autonomous operating weapons systems, of which some are still in an experimental phase and not subject to International Humanitarian Law.

The Unit of the Future

In light of the above scenario, considerable changes to the structure, command & control procedures and organization of the army are necessary. HCSS is not in the position to prescribe how these changes should take place and to which end-state, but we can assist in drawing focus to the areas to change and the direction of change:

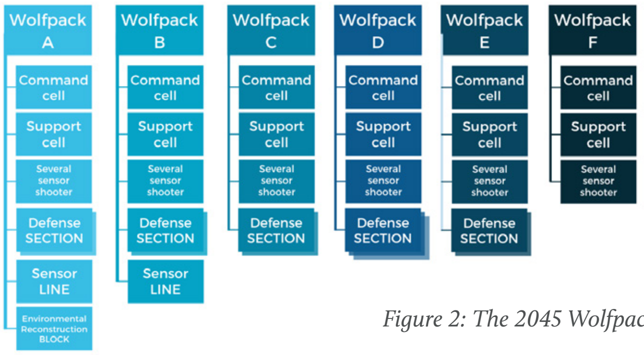


Figure 2: The 2045 Wolfpacks

Back to 2020: Recommendations towards 2035 and 2045

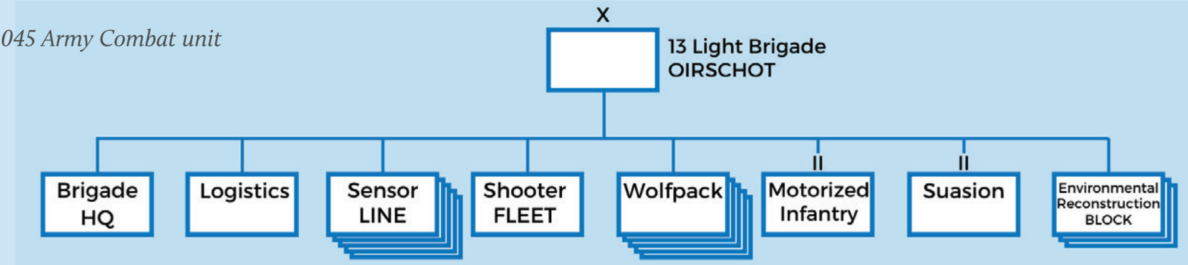
From 2020 towards 2035

- Stimulate leadership regarding fully integrated RAS as the future to prepare for
- Continue experimenting, including operations with RAS on all levels
- Be transparent to the public on ambitions and risks and (ethical) issues
- Engage outside world in confidence building and forming a realistic vision on RAS' future
- Develop appropriate policies throughout government

Towards 2045

- Create development teams together with industry
- Organize flexible pools of people from industry
- Be aware of being permanently forward deployed

Figure 1: a 2045 Army Combat unit



The Hague Centre for Strategic Studies

The Hague Centre
for Strategic Studies
Lange Voorhout 1
2514 EA The Hague
The Netherlands
hcss.nl

Chapter 5

The Implementation of Robotic and Autonomous Systems: The Future is Now, prepare for 2045

Patrick Bolder, Michel Rademaker and Bianca Torossian

1. Introduction

Over the last two years, HCSS has conducted research on Robotic and Autonomous Systems (RAS) in a military context concerning several aspects and dilemmas. Throughout this paper on the implementation of RAS, we hope to inspire thinking and stimulate the reader to reflect on the future use of RAS, draw recommendations towards the year 2035 that fit within the 'Operationeel Kader voor het Landoptreden' (and align these recommendations with the foreseen 'Defensievisie 2035') and consider recommendations for the implementation of RAS towards the year 2045.

The rationale behind looking far into the future is twofold. First, significant questions must be addressed early in the development and implementation of RAS. Many technologies are still in their infancy and similarly, our understanding of the political, strategic, tactical and, operational application of RAS is in its early stages. The second reason stems from the idea that people tend to overestimate the maturation of technologies in the short term and underestimate the speed of technological developments in the long term. Thus, by using both shorter- and long-term time horizons, room is created to think out-of-the-box whilst simultaneously lending opportunity to plan against a plausible, but—not yet ready—future.

This paper assesses some relevant elements for the implementation of RAS into the armed forces and especially the Army. It raises questions regarding the formulation of concepts and doctrines, how command & control over RAS is organized, and the consequences of these changes for personnel (including their training), logistics, infrastructure, organizational processes, and leadership.

Throughout this paper, questions raised will not always be explicitly answered. This is because, in many cases, it is still too early to provide clear solutions. However, further disentanglement of the issues mentioned will help discussions and, eventually, decision-making. In many cases, answers can only be realized after first experiments are conducted and experiences working with RAS are gained. Therefore, it is improbable that the thinking reflected in this paper is complete. The paper rather provides first thoughts and some conceptual points of view on issues the military will face in applying RAS within their organization and work. For insight into the challenges outlined in this paper, HCSS conducted an expert session using a serious game, the results of which are fully integrated into this paper.

Against this background, this paper develops recommendations regarding which lines of development or policies must be developed, the timeframe by which this should occur, and the prerequisites for these policies.

2. Methodology

Based on research HCSS conducted throughout the two-year project, Robotic and Autonomous Systems in a Military Context, elaborations took place regarding several dilemmas and issues pertinent to the implementation of RAS. During the project, papers have been developed on the operational applications, ethical dilemmas, legal aspects, collaboration and concept development and experimentation (CD&E).

As the first step in this paper, a future perspective is constructed through an anticipated plausible scenario and a short statement or point of departure for the analyses. This scenario is situated in the year 2045. The breakdown of the scenario into snippets of personal stories was chosen to make it as concrete as possible, and to ensure it was relatable and understandable for different audiences. Integrated within all snippets are the dilemmas or issues that were identified throughout the two-year research period. In the scenarios section (and subsequently throughout this paper) new terms, such as 'wolfpack', 'fleet' and 'line', are introduced to describe military levels of command. This enables us to detach from traditional thinking on levels of command.

A second step of the analysis conducted was the construction of a future RAS unit. Insights gained from the research paper on the military applicability of RAS were taken into account and then extrapolated into a plausible set of future systems. To build the future RAS unit, these systems were combined to form a unit that is only slightly akin to a current-day combat brigade.

A third step delineates four lines of development that explain in some detail what is required to undertake the implementation of RAS into the armed forces. At this point, HCSS conducted an expert session using serious gaming tools to gather further insights and validate our thinking.

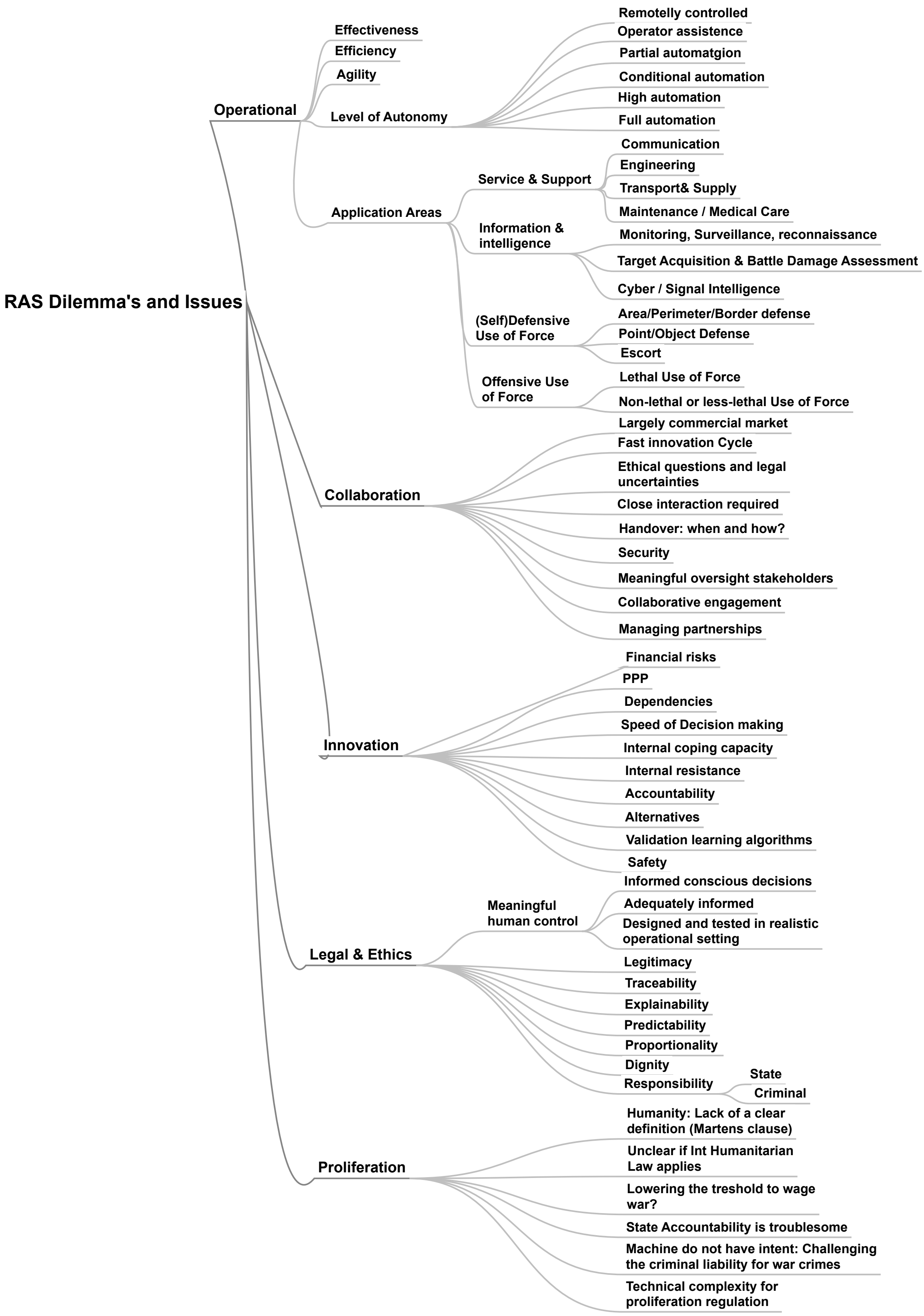

RAS Dilemma's and Issues
Operational
Effectiveness
Efficiency
Agility
Level of Autonomy
Remotelly controlled
Operator assistence
Partial automatgion
Conditional automation
High automation
Full automation
Application Areas
Service & Support
Communication
Engineering
Transport& Supply
Maintenance / Medical Care
Information & intelligence
Monitoring, Surveillance, reconnaissance
Target Acquisition & Battle Damage Assessment
Cyber / Signal Intelligence
(Self)Defensive Use of Force
Area/Perimeter/Border defense
Point/Object Defense
Escort
Offensive Use of Force
Lethal Use of Force
Non-lethal or less-lethal Use of Force
Collaboration
Largely commercial market
Fast innovation Cycle
Ethical questions and legal uncertainties
Close interaction required
Handover: when and how?
Security
Meaningful oversight stakeholders
Collaborative engagement
Managing partnerships
Innovation
Financial risks
PPP
Dependencies
Speed of Decision making
Internal coping capacity
Internal resistance
Accountability
Alternatives
Validation learning algorithms
Safety
Legal & Ethics
Meaningful human control
Informed conscious decisions
Adequately informed
Designed and tested in realistic operational setting
Legitimacy
Traceability
Explainability
Predictability
Proportionality
Dignity
Responsibility
State
Criminal
Proliferation
Humanity: Lack of a clear definition (Martens clause)
Unclear if Int Humanitarian Law applies
Lowering the treshold to wage war?
State Accountability is troublesome
Machine do not have intent: Challenging the criminal liability for war crimes
Technical complexity for proliferation regulation

Figure 1. RAS Dilemma's and Issues relevant during implementation.

3. Scenario: “2045: the era of relentless competition”

The Great Power Competition which started in the late 2010s has exacerbated over the last 25 years, never leading to wars on a worldwide scale, but instead resulting in high tensions globally.

3.1 Point of Departure

Due to geopolitical developments and continuous support from the Netherlands to the EU mission portfolio, a reorganization and restructuring of defense organizations has taken place. Deployment of units is no longer an aberration of peacetime training and education, but rather the new norm for Army units that operate on the ground. *Operation Permanence* has become the new normal whereby rotations from barracks to deployment areas are continuously taking place, but the footprint in theatre is as small as required, made possible by intelligent communications systems and innovative maintenance procedures. Reachback is structured in a new way so that those working in the Reachback offices are part of the actual operation without physical presence in theatre.

Work at the Reachback HQ involves creating the digital twin of the actual battlefield/area of operation. Data fusion from all available sensors in traditional military domains, as well as from the Electromagnetic spectrum, Cyber domain, and Cognitive dimension is delivered and presented by the Data Science Cell. Through an enormous amount of slightly different scenarios based on this data and its applications to the current operation, so-called Ensemble Comparison points out the most likely scenarios and their implications. These are then further analyzed to prepare the deployed units for their tasks. Through these means, it is possible to continuously observe the overall situation, learn from the theater digital situation overview and drill down on the specific critical localized elements for operational and tactical execution of tasks. Through targeted deeper intelligence gathering, (mostly automatically generated requests for information (RFI’s) by AI systems) we can now discern the needles within the haystack and can act upon neutralizing or utilizing those needles.

Other work at the Reachback HQ concerns the development and training of algorithms to feed into the deployed RAS. The continuously changing requirements for new algorithms are written here and the development of existing algorithms can occur in a Company-Owned Military-Operated (COMO) construct.

3.2 A short history of main events

3.2.1 Climate change

Climate change has had a more severe impact on world stability than the Covid-19 pandemic of 2020/2021. Competition between states and blocs established themselves more prominently under the pressure of a changing climate, leading to pressure concerning the security of supply of rare earth elements, agricultural areas, food, and water.

3.2.2 Geopolitical developments

Climate change and increasing national and regional self-assurance have led to a state of continuous competition, keeping world leaders on their toes. Cooperation is defined as ‘friendly but cautious’ between the powerbrokers, whilst at the frayed edges of the world, unsettling situations and events take place which need to be confined and managed to prevent spill-over to the more developed world. Even though there are regular meetings of the leaders of the BIG8 (North America, South America, EU, Russia, China, Australia, African Union, and the Group of Non-Aligned Nations), trust is not common amongst them. Outlier states (e.g., North Korea, Iran, Venezuela, India, and Pakistan) derive their position from unpredictable and erratic behavior, often supported by their possession of nuclear equipment and arms.

3.2.3 Consequences for the EU

It has taken time for the European nations to come to the same conclusion, but all saw their future best served by strengthening the EU and put to rest the continued disagreements about budget rules and national deficits. European regional autonomy is widely accepted by EU leaders as the only sensible way forward. In 2025, an ambitious EU policy was implemented, spanning more areas for communitarian cooperation than ever.

Security and safety nowadays are newly defined concepts and have a wider reach than they had in the early years of the 21st century. Former ‘basic human rights’ have become so pressured in other power blocs that they have been adopted in the EU as culture: morals and ethics worth fighting for. The newly defined notions of security and safety are not only about territory and free trade but also concern subjects as food security, human rights, development goals and freedom of gender, speech, religion, thought and expression.

As the EU’s economic strength is also served by global trade, border security had to be redefined as well. This was not achieved by completely closing borders but rather by controlling them to allow for seamless but safe flows in all domains as critical and innovative solutions were developed. These solutions were deployed not only in the physical domain but also in the cyber domain, which was much in need of advanced protection, rules, and laws. Since the space domain has become more accessible, critical materials are now also mined on nearby asteroids, which led to new kinds of borders with new kinds of security issues in new domains.

3.3 Main missions for the EU

The EU's posture is one of constant alert, leading to the continuous deployment of troops and their equipment along the sea-, air-, and land borders of the Union. The cyber domain is an a-typical phenomenon as borders are not easy to discern, but security here has become almost more adamant than in the physical domains. Also, security in, to, and from the space domain is now one of the missions the EU Security Council decided upon in 2042 when the new EU Security and Safety Strategy (EUSSS42) was adopted. Within the European security environment, burden sharing, and mutual responsibility became the new norm for the enhanced and intelligent border protection missions. For almost all European nations this means that military units are on a rotational schedule for the basic set of the so-called 'Primary EUGuard Missions' (PEM) as described in the EUSSS42. These vary from '*space surveillance and evasive maneuvers*' to '*maritime patrol and engage*', '*air defense and deter*', '*land observe and deny*' and '*cyber protect and defeat*' missions. The required stamina for these missions could not be guaranteed by human force alone; developments in AI and Robotic and Autonomous Systems (RAS) were not only welcome but instrumental in safeguarding the European continent.

3.4 Important role for the Netherlands

The Royal Netherlands Army had already started to experiment with the deployment of robotic systems in the late 2010's. High numbers of RAS make up for the lack of sufficient personnel. This enables the Royal Netherlands Army to execute prolonged forward deployment for the *'land observe and deny'* PEM-missions along Europe's stretched borders. As the Netherlands (with The Hague as the City of Peace and Justice) is known for its application of law in all facets, it is seen as the guiding country concerning autonomous unmanned systems and its implications for all aspects of legal and ethical issues such as (amongst others) 'meaningful human control' and legitimacy in the use of violence.

For that reason, the Royal Netherlands Army is currently heavily engaged in the '*land observe and deny*' EU-mission called OSIRIS (Operation Southern International Reaction-Intervention-Security) in the northern Nile delta in Egypt. Their mandate comprises border security and prevention from human trafficking, weapons smuggling, and data crimes.

The challenges for the Royal Netherlands Army unit on this mission are considerable. The opponent, the Free Organization of Egyptian Identity (FOEI), supported by troops from neighboring countries, such as Sudanese fighters, has acquired advanced technologies in unmanned systems and Artificial Intelligence. There are strong indications that the Russian-affiliated Beethoven group is heavily backing FOEI forces. Though Moscow denies involvement, international passenger records show that several young Russian men and women are having holidays in the region. FOEI leaders have said that they rely on the force and fearlessness of autonomous weapons systems, even if some aspects of them are still in an experimental phase and not made subject to International Humanitarian Law.

3.5 What does the future look like from the perspective of a user?

In these snippets of the future, both our own troops and adversaries come into play. We have sketched their roles and have included dilemmas and issues, addressing them in a plausible future where RAS applications might have a prominent role to fulfil.

EU-OSIRIS:

Camilla Draper, CEO of Draper Robotics had a short night. After having been up for some 24 hours, she was again awake very early in the morning. Last night, her company was requested to update some of the applications of the unmanned systems running in the Northern parts of the Nile river borders. Dutch land forces are participating, as part of the EU-mission in this hostile territory and some of her employees have a part-time reservist role embedded in the unit there too. Her contracts provide what is called Company Owned Military Operated (COMO) services. She provides on short lease drones and autonomous ground-based robots and even two types of amphibious and underwater surveillance drones. The military uses these robotic systems, and her team assists in the upkeep and technical operational services both back home and in theatre. For some, logistic functions Draper Robotics provides full services, meaning that they do the maintenance, keep stock, have SLA's for providing replacements of modules or complete systems within max six hours from in-theatre, dispersed forward storage and maintenance sites, etc. These sites are integrated into the logistics information systems the military units use. For use of force the military are the operators. But it is the combination that made her company stand out. For her, it is profitable and for the military, the benefit is flexible and diverse capabilities, cost-effectiveness and the ability to focus on core-tasks. She managed to update the unmanned systems within 30 hours, implementing the software her team developed over the last two weeks. Even after this short night, she could not sleep anymore as she realized that she was on a slippery path now regarding the handover of her new software to the military, as the division of liability between the operators and her company became questionable by this update.

FOEI troops and its affiliates:

Boris Krygizie, director of BEAR Robotics, had an exhausting week. For seven days in a row, he has met the chief innovation of the FOEI military RAS unit and the Army staff section that oversees all commercial parties during the life cycle of the unmanned systems. They discussed at length the latest system feature that would allow one of the systems to threaten and/or harm a human being while interrogating. According to Boris, this innovation could help the forces to win the conflict in the Northern parts of the Nile river borders against the EU-mission. However, according to the Army staff this might be a tactical win but could endanger the legality of their mission in the eyes of domestic and international society. In the end, all his hard work went down the drain as it was decided that the latest feature would not be installed. However, all is not lost: Boris might sell the feature to a friend who is the director of a private military company that operates in Mexico.

EU-OSIRIS:

LtCol Jack Jansen has become used to COMO-services over the last ten years and has seen them mature during his time commanding three WOLFPACKS[1] in the Nile Delta theatre. They have a soldier-to-robot ratio of 1:6. He has 75 troops of which 50 are partaking in 24/7 operational combat and combat support tasks and up to 300 robotic and autonomous systems (RAS) to help them in both defensive and offensive roles. This is the equivalent to the capabilities of a battalion back in 2020. Most of the RAS are integrated, meaning that they are partly coordinated autonomously and work in swarms up to 20 (a number tested as being effective for a single operator to handle at this stage of development). Every combat soldier can control a swarm and extend their capability tenfold. During several missions, it appeared that his unit had a high deterrence posture. A WOLFPACK is nowadays capable of operating at two levels at the same time. Jack walks alongside conditional automated systems that defend the perimeter. Though the systems are switched from 'conditional automated' to 'operator assisted' meaning the systems will not fire on him autonomously, Jack, despite so many years of experience with RAS, is still not very comfortable walking in front of the autonomous systems.

As RAS provide full situational understanding and can integrate vast information across the operational theatre and at home, unused capacity is automatically identified, and the surplus is available to his unit (and vice versa). With drones and unmanned systems, the mobility of his unit has improved dramatically, and the footprint has reduced. In addition, his effective warfighting reach for the smaller systems has extended from an average of 3 kilometers around his unit elements to some 30 kilometers, meaning that his coverage of theatre is six times larger than it used to be. For him, as a commander, it gives a much better awareness and understanding of his options. The situational picture has greater depth and updates are constantly available, which means that his briefs and operational orders have a different nature. The situation at hand is sketched automatically and several plausible scenarios are generated including plausible courses of action. Because of the availability of vast data from all sensors in his unit and others, data analytics supported with AI provide a better risk assessment than he was used to some ten years ago. Simulation runs, or so-called "dress rehearsals" are the standard practice these days. But all-in-all his skills mastering the art of war are still required.

FOEI troops and its affiliates:

Former LtCol Youri Nikolajev has been heavily involved in RAS units over the last ten years. He operates in the Nile delta as an advisor to the FOEI militia that controls several unmanned systems. Some of these systems can be adapted to the situation and the task within 24 hours: one day the system defends in high automated modus an object, and the next day the system conducts an attack on the EU soldiers in remotely controlled modus. Youri analyzes the options

[1] Throughout this section, and subsequently throughout this paper, new terms, such as 'wolfpack', 'fleet' and 'line', are introduced describe military levels of command. This enables us to detach from traditional military organizational thinking and broaden our creativity, and furthermore represents how our conceptual framing of units will shift in the coming decades.

his Command & Control device has produced for the upcoming operation next week; now he must decide whether to use his less sophisticated drones for communication, for supply, or medical care. These legacy drones require 5 days of rebuilding and software-updating. Luckily, he has received plenty of modern information and intelligence systems in the last month. These systems provide him with a clear picture of the EU forces. Besides that, these systems propose which targets to engage. Youri is convinced that at the end of next week his Egyptian militia will have won a major battle. It could have been better if the militia had not wasted the five mini-mine laying systems two weeks ago.

EU-OSIRIS:

Major Estella Hansen has recently changed her position from project-leader in the Defence Material Organization to commander of an operational RAS unit. She trains her operators to understand how their systems 'think'. According to her, the operators must understand how their systems decide to engage targets because in the end the people behind the systems are accountable for the behavior of RAS. If the operators are not able to predict the system's behavior to a certain degree, they are not able to override the system if it malfunctions. She knows that her systems do not decide well on proportionality. Therefore, the operators must override the systems during offensive operations using lethal force against the Nile delta militias to avoid civilian casualties and stay within International Humanitarian Law.

FOEI troops and its affiliates:

Major Umit Sjukoev has recently changed position, from the commander of an operational RAS unit to project-leader in the Airforce staff's RAS section. Now he is in the position to oversee the whole life cycle of his beloved unmanned systems. At this moment, Umit writes the requirements for a new drone. The drone must decide itself which target it will attack. Chapter 3 of the requirements describes how the Airforce keeps an overview concerning the drone's adherence to the ethical regulations set by his country from the designing phase through to the manufacturing, testing and operational phases, and finally the decommissioning phase. The telephone rings and Umit must explain once again that the systems, contrary to human beings, do not get tired and emotions do not influence decisions.

EU-OSIRIS:

Captain Jan Jager directs the Lethal Autonomous Weapon Systems towards the Sudanese militia that marches in large columns to their defensive positions close to the Nile. Despite his relative weakness (due to the small number of systems) he was not afraid to attack this massive militia nor does he care about losing his systems, as he will receive another batch next week.

One of Jan's concerns is to adhere to the principles of International Humanitarian Law. For him this is very difficult because RAS are not mentioned in International Humanitarian Law, but Jan intends to follow the principles embedded in the tailor-made rules of engagement. The time he has used to come to the decision to attack has lasted much longer than the attack itself will take. He believes that an outsider could think that such a decision to wage war by RAS against humans would be easy but for him it was not.

Within a few minutes he is about to let his systems go. As of that moment, the systems are on their own, taking decisions to destroy, based on previously set rule-based engagements. Jan knows that some mistakes the systems will make will not be traced back to him or the system's manufacturer. He has no idea which entity would be accountable in such a case. He pushes the button to let his systems attack the militia.

FOEI troops and its affiliates:

Militia leader Abdo Majok directs his massive militia in marching columns to their defensive positions close to the Nile. Abdo expects an attack by the EU unmanned systems units on his left flank. He doesn't mind losing half of his militia as within three weeks he will receive new soldiers. He has tasked his left flank security patrol to capture some of the enemy system, he aims to sell these captured systems to terrorists in Nigeria. For Abdo it was easy waging war to gain money. He just must be careful that his government was not accountable for the activities of the mercenaries in his government-owned militia.

EU-OSIRIS:

Colonel Mats Verbraak has been preparing the implementation of the COMO-contracts for RAS since 2025. The land forces underwent a steep learning curve, as did the companies that provided the COMO-RAS-services. The 'teeth-to-tail' ratio has been tremendously improved. Training and exercising are more efficient and effective, and the speed of innovation has more than tripled. Much of the training is provided using fully automated simulations not only in the preparation phase but also in theatre. All these capabilities are provided in close collaboration with knowledge-intensive partners and COMO-companies. Due to the uncertainty regarding the amount of effort it would take to develop and test an innovation, the Army staff is flexible with his budgets. Mats appreciates this and continuously informs the Army staff on the financial aspects of his COMO-RAS-services. His strategy is to get at least two manufacturers involved in one system to facilitate competition between these manufacturers, otherwise he would run out of his flexible budgets. Mats' headache worsens as he must decide whether to incorporate the sensor-module in the next phase of the drone or in a later phase. Neither the sensor-module nor the drone is mature enough now, but both might be in time. Uncertainty and risks further complicate Matt's dilemma, but he must decide quickly. If he incorporates it in the next phase, he must use the scarce testing capacity from his other RAS project.

FOEI troops and its affiliates:

Colonel Bukin Sarachov has been involved in the implementation of RAS since 2025. Normally, he likes the speed of innovation with all the related challenges. But now, he prepares for the hearing with the General Court of Audit: he must explain the failed implementation of an autonomous transporter for wounded personnel.

His decision to speed up the testing phase required regular units to change their schedule of exercises. To achieve this, he fought internal bureaucracy to start the testing phase within 4 months. All the tests failed; the whole project turned out to be a disaster. The human-centric units blamed him for wasting so much money on useless innovations. Bukin will explain to the Court that he had pushed the testing phase too early, but there were alternatives less autonomous than the systems that failed, that could have been procured quickly. However, due to the widely spread criticism, it was politically impossible procure an alternative system. Bukin will elaborate in court a previous experience without a proper testing phase when he fielded an ammunition supply system in the Nile delta two years ago. The system's algorithms learned a lot and the system developed into a more autonomous system after a year. In that case he received the critique that the algorithms were not validated enough because learned behavior had been gained in an operational environment instead of in a controlled testing environment. Indeed, in a few cases these autonomous supply systems failed to deliver the ammunition safely, however, not everyone is willing to bear responsibility for this issue. Walking into the court's largest room, Bukin noted that some colleagues were not as open-minded to RAS.

4. The Unit of the Future

Considering these fictious—but foreseeable—scenarios, considerable changes to the structure, command and control procedures, and organization of the army are necessary. This includes all kinds of Tactics, Techniques and Procedures (TTP's) and perhaps doctrines. HCSS is not in the position to prescribe **how** these changes should take place and **to which end-state**, but we can assist in drawing focus to the **areas to change** and the **direction** of that change.

In order to give a hands-on idea of what a future unit could look like when RAS would be fully applied, a plausible unit was constructed. The insights gained from the research paper on the operational applications of RAS were considered and extrapolated into a plausible future unit that is only slightly recognizable comparing it with the current-day 13Th Brigade.

In 2045, the organization of the 13th Brigade is focused on RAS developments. The organization does not encompass all developments in other expert-lines (such as logistics or cyber-operations for example) up to 2045. The scenario outlines how the brigade is deployed in the Nile delta, with some elements more detailed than others. The non-deployed ('peacetime' organization in the Netherlands) organization might be different from the deployed one.

1. Brigade HQ

a) Staff are static in the Netherlands, the Reachback HQ

b) Staff forward deployed

2. Logistics

a) Medical company

b) Software repair, development & test company (including Reachback group to Army RAS organic capability and to the companies that provide the 'Company Owned Military Operated' services)

c) Hardware repair, development & test company (partly manned by Camilla Drapers civilian technicians)

d) Supply company (a multi-UGS and UAS in a network)

e) Robot recovery company

3. 5 LINES of sensors

(each LINE consists of an Analysis Cell with several sensors, UGS and UAS)

a) The analysis cell is split in a forward deployed element and an element in reach back in NLD.

b) The non-disposable UGS is the Cyclops RAS which conduct surveillance, target acquisition and reconnaissance. With its on-board AI-fueled systems it churns out intelligence products. Where essential, to prevent human decision-making latency, self-defense systems are employed by the Cyclops.

Smaller systems that can operate in urban areas, in wooded areas, etc. belong to the inventory. Further on, each LINE of sensors possesses a huge amount of small disposable UASs that are able to operate together with loitering munitions of the FLEET.

4. 1 FLEET of shooters that covers all areas of the Brigade.

It employs small to large munitions in massive amounts, partly loitering munition (short and long endurance), partly the munitions are able to attack in swarms. The hardware consists amongst others of Israeli developed, Dutch (REKKOF Military Industry) produced Stingray Multi Area Target Suppression (Stingray SMATS) Systems.

The loitering munitions are able to operate in swarms with the LINES' small disposable UASs. These swarms can mount up to 5500 UASs and loitering munitions.

5. 6 WOLFPACKS of sensor-shooter-combinations

a) Command cell

b) Support cell (including software development group, civil engineers of COMO services, Forward cyber operators, Forward non-lethal influencing operators, Electronic Warfare group, logistics) supply/maintenance/medical)

c) Several sensor-shooter-combinations (UAS and UGS) equipped with:

i) The standard Wide Range Observe, Precision Application Fire (WIROPAF) Unmanned Ground Systems for automatic close-in 450 degrees (360+90) able to destroy or suppress a target in lethal and non-lethal ways.

ii) The short range UAS 'Observe and Fight Bird' can be equipped (for each action) with a certain ammunition, combined with a couple of sensors.

6. 30 Defensive SECTIONS; defense against ground, air, and electro-magnetic attacks,

meant to allocate to other Brigade actors for self-defense of that actor.

a) Command cell

b) 22 soldiers with 50 ground defense systems

c) 18 soldiers with 40 air defense systems

d) 13 soldiers with 10 electro-magnetic defense systems

Consisting of the family of Hornet's Nest unmanned systems, due to its easily understandable and accessible AI applicable in multiple domains and easy reprogrammable and adjustable for various tasks.

7. **1 battalion of motorized infantry** (3 companies of motorized infantry with each 3 platoons)

   The platoons use Milrem Robotics latest UGS for a wide area of support and relief tasks enhancing the soldiers' effectiveness in battle.

8. **Suasion battalion**

   Electronic Warfare, cyber operations, non-lethal-behavior-influencing-capability

9. **4 Environmental reconstruction BLOCKS** (Mine laying systems, Demolition systems, Mine clearing systems, Bridge laying systems, Breach systems)

   Equipped with highly technical advanced UGSs where, based on AI and recognition algorithms, largely autonomous activities can be delegated to. The systems can be remotely operated via a datalink and ground control station by an engineer operator. The mine laying and demolition systems can also be tasked by the obstacle plan enhanced by the current 3D photomap for autonomous task execution.

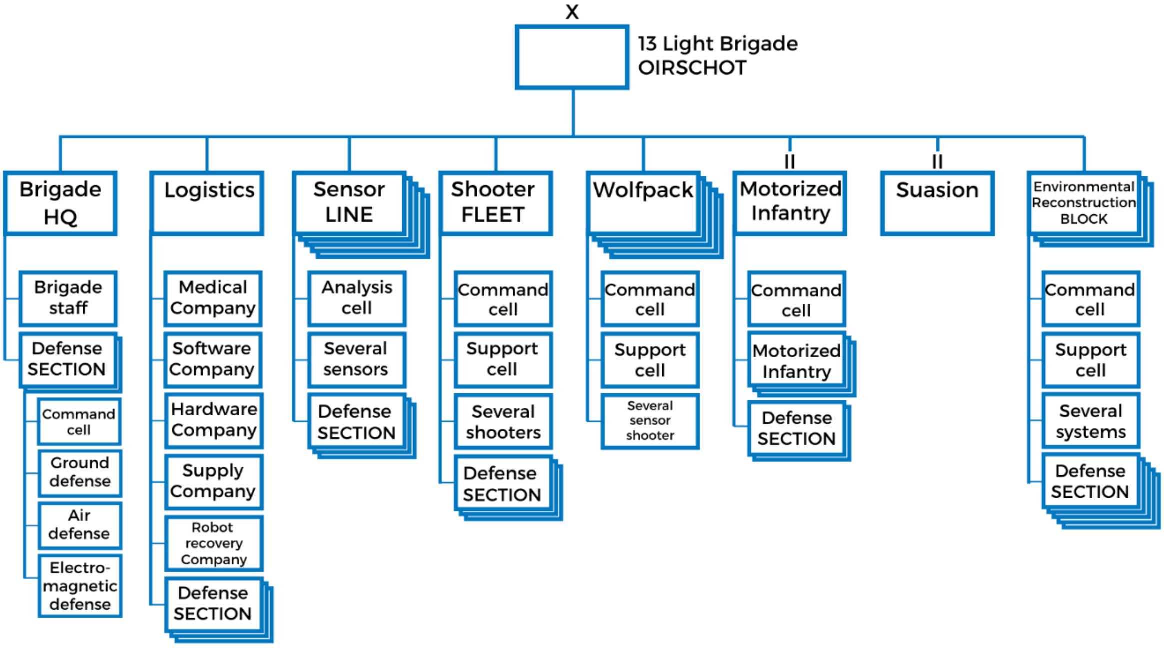


Figure 2. A 2045 Army Combat unit

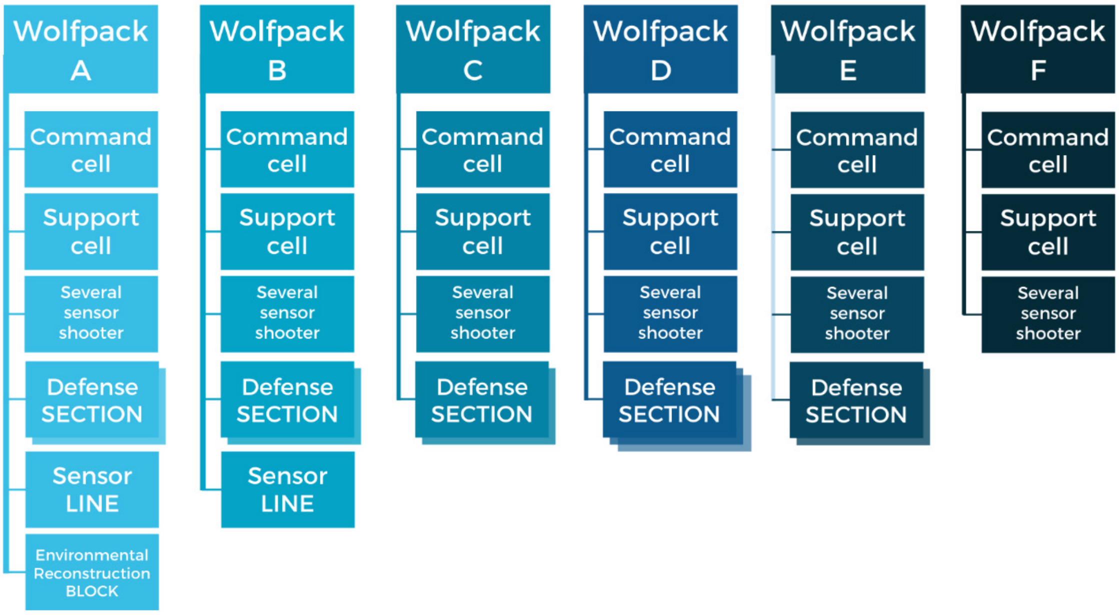


Figure 3. The 2045 Wolfpacks

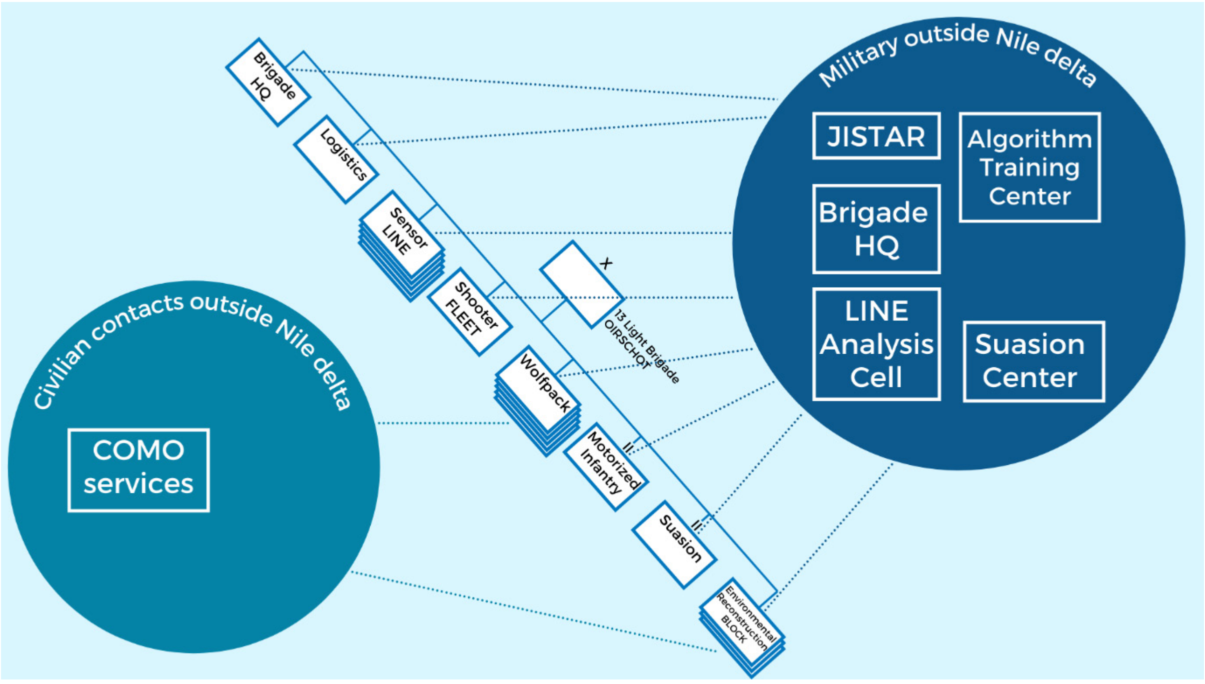

Military outside Nile delta
JISTAR
Algorithm Training Center
Brigade HQ
LINE Analysis Cell
Suasion Center
Civilian contacts outside Nile delta
COMO services
Brigade HQ
Logistics
Sensor LINE
Shooter FLEET
X
13 Light Brigade OIRSCHOT
Wolfpack
II
Motorized Infantry
II
Suasion
Environmental Reconstruction BLOCK

Figure 4. The 2045 Reach Back construct

5. Back to 2020: Lines of Development

For the implementation of future RAS capabilities, various initiatives must be undertaken. Based on the plausible future scenarios explored in the previous section, which represented many of the dilemmas and issues at hand, a series of lines of development can be identified. For each line of development, a description is outlined, and relevant aspects are addressed and operationalized by linking items with policy developments and other activities already underway. This includes strategy and plan development, policy and operations concept development, innovation, adaptation issues, finance, recruitment and human resource issues, norms, legal queries and public support.

The description for each Line of Development lists philosophical and open-ended questions which are designed to provoke in-depth discussions. These can relate to new doctrines, concepts, working cooperation, business models, etc. It is foreseen that developments will continue to take place and constant adaptation of the organization is required. The biggest lesson to be learned here is that to tackling all developments in a coherent way requires leadership and trust.

5.1 Line 1: Development and acquisition

Technological developments are underway but still require a significant testing.

We are at the stage when (r)evolutionary technical developments no longer stem solely from military research and developments. Long gone is the period when advanced military technology was introduced for soldiering purposes, eventually finding its way to civil society for day-to-day peaceful and domestic use. Commercial companies, producers and factories nowadays see innovation as a means of survival, and it is thus embraced with much enthusiasm. This attitude towards thinking about constant renewal of business processes and products takes place throughout the commercial world. The civil techniques and novelties found there can fulfill military tasks after some adaptation if necessary. However, for purely military tasks (e.g. (supporting) warfighting, civil-developed innovative techniques) RAS will probably need considerable adjustment to be effective. From this perspective, it is fair to say that in the field of military RAS, civil-developed techniques will have to be reinforced with military knowledge and specific development. At this point, RAS and especially military RAS, are still in its embryonic stages, meaning not much is tested and ready for action, let alone ready for immediate use.

Here arises an important decision point: Should the Netherlands armed forces act as a smart buyer of existing technology, or do they place themselves at the forefront of technological advancement and be a part of the design and development? As previously stated, RAS is in its first stages of development and if the Netherlands Armed Forces do not have the luxury to wait for military off the shelf products, it will need to engage in the development of RAS. RAS will probably have to be developed according to specific demands from the MoD. For the Netherlands, this represents an excellent opportunity for triple helix cooperation whereby knowledge institutions, commercial firms and factories, and the armed forces can articulate the need and applications for (different types of) RAS, set design parameters on hard- and software, and undertake prototyping activities. Here lies the best chances to provide specific military needs. In a form of spiral development, where all parties concerned work close together and have short feedback loops, such an approach offers the best chances for quick success.

The development of RAS will probably occur in revolutionary steps as through each phase of advancement new areas and possibilities will be discovered and defined where RAS can serve and add value to human skills or even replace them. To foster such a development path, serious thinking should be allocated to realistic, yet safe test environments. The result of a spiral or (r)evolutionary project-design should consist of a technology demonstrator, leading to an operationally usable and effective prototype. This prototype will likely lead to a better understanding of capacities and lead to further ideas and idea expansion. The above-indicated approach fits within the Defense Industry Strategy recently adopted by Parliament. The Ministries of Defense and Economic Affairs are strong supporters of providing domestic industry with the best chances to participate strongly to the profitable defense and security market. Such a movement requires an attitude of entrepreneurship within the Defense organization in order to bring all involved parties on board. But due to internal sets of rules and regulations, even when this point is reached, it will also lead to challenges in the field of acquisition and procurement. A culture where innovation is fostered as a competency of the utmost importance will contribute to making the above developments a reality.

Current guidelines on defense acquisition have firm rules regarding business competition and offering a level playing field to all interested industries. However, it is sometimes difficult to involve ‘first-stage developers’ who are in the stage of offering their product into competition with other possible providers. Due to their previous involvement, they are in possession of important and strategic knowledge which could give them an advanced position and consequently would deprive other suppliers of a fair competing chance. The result can be that the company which bares the biggest financial and technical risks is excluded from making a profit that reflects their degree of risk. By publicizing their findings to the market, a level playing field is created, but companies lose their knowledge position as a result. This is a significant deterrent for adventurous and innovative firms to participate in the development stages of new weaponry. Subsequently, it can be reasoned that such an approach halts innovation. Newer approaches on acquiring technologies are leaving behind the buying new systems. Leasing and using the capability without owning the system itself, is a business model that will become increasingly more present in the military. The distinction between ‘owning’ and ‘using’ will have to be explored, particularly when it comes to responsibility for maintenance and malfunctions.

Another issue to tackle within procurement policy is in relation to numbers and batches. The military prefer to have a high degree (if not the maximum) of commonality throughout a ‘fleet’ of certain weapon systems. In their eyes, this will ease the logistic and maintenance efforts because the production and servicing of the equipment can occur *en masse*. Further, in terms of interoperability, there are advantages of a common fleet. With RAS, development will never be over and improvements and enhancements, as well as new capabilities, will be added over time. Instead of looking at maintenance issues simplified by standardization, the positive side of maximum capabilities and possibilities should outweigh the perceived negative ones.

Innovation partners

The construct depicted in the partial scenarios is on COMO (Commercial Owned Military Operated), a business set-up that has no preceding examples when it comes to the use of military equipment in active firing zones. This will challenge the creativity of contract managers and will lead to the development of advanced business contracts. The innovation of processes and capabilities will have to take place within the Defense organization and new avenues of approach will be invented. As this is about equipment that will not be completely out of the development stage for the coming years, the innovation partners should not only include the hardware and software providers, but also knowledge institutions and other non-standard scientific disciplines.

5.2 Line 2: Operational Excellence

Redesign of concepts, capability packages, doctrines and TTP’s will require a reshuffle of capabilities when developing and implementing RAS. Traditional capabilities might be obsolete or less effective, while others might be boosted even more.

Military operations, their planning, tactics, and conduct have their roots in concepts and doctrine. Concepts and doctrine describe what is done in operations, how they are done and which rules to act upon during operations. Concepts and doctrines prepare soldiers for the ‘fog of war’ during operations. Circumstances and conditions change during battle, which are sometimes or partially foreseen. Training for all exceptions and aberrations from the set-up plan is impossible but preparations towards the unknown and unexpected can and must be done by utilizing concepts and doctrines. Concepts and especially doctrine will give soldiers confidence through which they can fulfill their tasks even under harsh and dangerous circumstances. Concepts and doctrines in this way also define training and preparedness. Though concepts and doctrine fulfill purposes and are often the backbone of planning and training for operations, they are not set in stone and should be subject to evolution over time and practice.

The introduction of new systems can lead to the performance of new tasks or a change in the way existing tasks are performed. This is especially the case when these new systems are revolutionary in nature rather than a next iteration of existing weaponry. In this case, new types of weapon systems can, and in some situations, must lead to the adaptation of concepts and doctrines. It is likely that over the years the value, the number of tasks, acceptance and usability of RAS in the military environment will expand. The first RAS will be assisting human soldiers in simpler tasks such as carrying heavy equipment, scouting ahead and setting up secure communications. Gradually, RAS will become assistants of human fighters, supplying and sending intel through their advanced audio and visual sensors, including through the infrared spectrum and radar, whilst simultaneously receiving information from other sources as well. Another task of early RAS can be maintaining radio contact with higher and lower echelons and neighboring units, ensuring that the supply lines for equipment and the transport of casualties from the battlefield is organized, all whilst being accompanied by a precise GPS-tracking system to prevent friendly fire. RAS will also be carrying heavy weapon-systems and possibly operating them alongside human soldiers. A further development will be circumstances whereby RAS replace human soldiers almost completely, especially in extremely volatile circumstances, unfit for humans. In all stages, the freedom of movement of the employed RAS will have to be defined and the level of autonomy decided.

All these different modes and levels of operation must be ‘learned’ by RAS, but also or even more so by the human operator or the human working alongside RAS. How this cooperation should take place, and the most effective means of man-machine teaming, will have to be developed and put into doctrine. Especially in cases where RAS will take over increasingly complex functions from humans, these doctrines are essential for optimum use of all capabilities that RAS offers and to ensure the operations are as safe and ethical as possible. Ever-tightening decision loops require increasingly quicker sequences of observation, orientation, decision and action (OODA-loop). One way to approach this accelerated decision-making is by introducing high levels of autonomy within weapon systems. For defensive weapon systems, this is probably the only way to be inside the opponents’ OODA-loop.

Introducing RAS within the military might ultimately lead to broad changes within organizations, as indicated earlier in the ‘Future Scenarios’ section. This trend of applying increasing levels of autonomy has already taken place. Armored vehicles like tanks and APCs can be equipped with reactive armor, chaff and flare can be dispensed in automatic mode from endangered aircrafts and naval close-in self-defense systems must be on automatic mode in order to be effective at all. Air defense systems already can apply a considerable amount of autonomy as proven by the Goalkeeper, Patriot and NASAMS systems. Once on automatic setting, they can detect and identify targets and decide to launch ammunitions and missiles to neutralize those targets without human interference. Although these systems have limited tasks, they have introduced a level of autonomy which has become acceptable. The further levels of autonomy will become not only acceptable but essential.

Another challenging doctrine development will take place when RAS are deployed in fighting missions and especially where they are at the forefront of the conflict. When humans are wounded the so-called golden hour is applied: within the hour the wounded must be treated in a hospital. If a soldier is killed, colleagues want to make sure they are not left behind on the battlefield. Do we apply the same type and level of ethics to machines? What rules and doctrine concerning damage will we have for RAS? If a system is deemed too damaged for operational use, what should be done with it? Should it be left behind because it is just another piece of machinery, or should it be made certain that RAS does not fall into enemy possession potentially contributing to technological spill and the destruction of our tactical advantage? Should the RAS self-destruct with the risk of hurting our own people or equipment? Which types of RAS can we leave behind when damaged (wounded) or destroyed (killed)? Does that answer depend on the level of autonomy, lethality, usability, tasks to be performed by RAS? Or does it stem from the technology installed in the RAS and if it can or cannot fall into the opponent's hands? All these questions are important to have answered before RAS is being used in training and operations. Clearly, the change in concepts and doctrine will move from human-centric operations assisted by RAS, towards optimizing man-machine teaming and eventually RAS-centric operations supported (and directed) by humans. The answers to all the above questions will form part of this doctrinal change.

Command & Control

Military operations are very much task- or mission-oriented, whereby individual soldiers must have the ability to perform the mission according to the commands he received from a higher authority. In an operational military environment, these commands will be given according to a certain set of rules, regulations, and vocabulary. At the same time, the soldier must be able to adapt to changing operational environmental circumstances and still reach the desired effect and complete the task. Even when—or especially when—he is unable to reach a higher level of command for guidance in these new circumstances. Therefore, while on the one hand, military operations require discipline to follow orders (a core military virtue), on the other hand flexibility of mind and creativity of the individual is required.

One might argue that concerning the required discipline on one side and the needed flexibility and creativity on the other, RAS will likely be difficult to integrate within the military on anything more than just 'dull', 'dirty' and 'dangerous' tasks. Indeed, this is where the first application of RAS within the military will take place. They will alleviate foot soldiers hard work (e.g., packbots) or take care of navigation, thus reducing the strain and fatigue on the soldiers and allowing them to be more focused on warfighting. Such a relatively simple task will already require special skills from RAS. They should be near, following the soldiers without hindering them and be as silent as possible. Escort systems must be able to hide and run when the soldiers do. RAS should be as independent as possible, operate with minimal commands, and yet be a reliable partner. It is debatable whether these skills require a form of AI installed in the RAS, or if smart programming will provide all needed capabilities and operating modes. But as indicated earlier in this paper, once RAS have proven their added value, the amount of RAS will increase as well as their tasks and utilities. Obviously, more complex tasks require more 'intelligence' in the RAS.

Humans and RAS excel at different cognitive tasks. Close attention should be given to how RAS will gradually take over human tasks, without degrading the number of considerations humans have in their decision-making and task execution. Especially in situations where human life is at risk and International Humanitarian Law comes into play, we are still reluctant to allow AI to make independent decisions although, in practice, AI is already widely used for relatively simple and harmless tasks. The operator for less autonomous RAS likely is in a safe environment (for example MQ-9 operators), has Reachback to all kinds of supportive systems and can confer with colleagues and leaders about decisions to take. In such a case it might improve the quality of these decisions and make them considerably more thought-through than those taken in the heat of battle. Furthermore, they do not get tired or distracted and can perform tedious tasks without losing concentration and focus. Decision-making here is based on algorithms and lines of programming, which in principle does not falter. When decisions taken by RAS are based on imagery and visual recognition techniques and patterns, there is no human 'filter' in place to make unsubstantiated deductions. When RAS can make use of deadly force, the ethical debate concerning Meaningful Human Control (MHC) comes into play.[2] Noteworthy is the discussion about Meaningful Human Control, Human responsibility, and Accountability.

Soldiers will use (deadly) violence in war according to a prior set of rules made often by politicians as the Rules of Engagement (ROE's). These guide the decision-making for the use of weapons without the need to ask for permission every time a soldier feels it necessary to use lethal force. Without reiterating the debate on MHC, one might argue that when ROE's are instilled in RAS software, and kept up to date, that this is also a form of MHC. ROE in the C2 system of RAS might become a way to guarantee HMC in autonomous systems. Command and control are also a two-way street. A command is given and received, the receiver should study the task given and report back if and how they will perform the task, what they need to complete it and sometimes what they expect from others to execute it. After the action, a report should be given to the higher echelon about completion of the task, mission effectiveness and peculiarities, if any. Another question arises on how C2 will be conducted in future when intelligent RAS will make up a serious amount of the military inventory. Thinking further ahead, will there be levels of command between RAS, are there any boss-RAS and subordinate ones, can RAS in certain circumstances 'command' humans? Basic throughout the use of RAS will have to be that they are relatively resilient to cyberattacks and thus for instance cannot be turned (by the opponent) against the employer of the RAS.

5.3 Line 3: Legal and Normative frameworks and Public Support

The application of RAS will be constraint by issues of ethics, proliferation, laws and regulations, public understanding, and support.

The use or potential use of RAS has been at the forefront of discussion in the public sphere. On one side of the debate, RAS are sometimes framed as 'killer robots' with the associated risks stemming from the combination of autonomy and weaponry are highlighted, thus narrowing debate to

[2] See Esther Chavannes and Amit Arkhipov-Goyal, "Towards Responsible Autonomy," The Ethics of Robotic and Autonomous Systems in a Military Context (The Hague: The Hague Centre for Strategic Studies, September 2019), https://hcss.nl/sites/default/files/files/reports/Towards%20Responsible%20Autonomy%20-%20The%20Ethics%20of%20RAS%20in%20a%20Military%20Context.pdf.

autonomous weapons systems and drawing away from the vast, non-lethal applications of RAS. The other side of the debate emphasizes the potentially critical role of RAS in gaining competitive advantage in conflicts during an era where the character of warfare is rapidly transforming. This complex debate requires deep thinking on (future) ethical implications but must be grounded in the reality of the conflict environment around us (and ahead of us) and the vast opportunities available for RAS implementation across numerous military application areas.

TNO's thinking on this and their development of algorithms such as Goal Function and World Model is a promising approach. RAS, for one thing, will not always be equipped with lethal or less than lethal weapon systems. In the case of the latter, will an ethical decision on employing such systems in a military operation be necessary or even be of any added value? Will such systems be subject to the Wassenaar agreement on arms sales? Will it be necessary to make these systems subject to the Convention on Certain types of Conventional Weapons, or are they weapon(systems) at all? And when RAS are armed (e.g., an MQ9 with Hellfire missiles) does this make it automatically into an autonomous killer drone as long as the decision to fire the missile is made by humans and still done based upon acquired intelligence? The fact that the weapon system itself acquires footage which forms the basis for the decision to apply force executed by the weapon system still does not make it a killer robot. For example, a modern Sidewinder missile fired by a pilot in an air-to-air conflict acquires its target after being cued towards it and follows it and deploys its autonomous sensor to decide the best point of impact or explosion as well as taking into account evasive maneuvers and tactics such as flares in order to achieve the desired effect. In the case of the Reaper Drone armed with Hellfires, it is nothing more than a very concise sensor-to-shooter loop, built especially for time sensitive targets. It seems that new norms, descriptions and taxonomies are required and should be established in order to structure valuable debates and to make the debates worthwhile when they concern the bigger issues such aa life or death decisions.

One might state that the developing knowledge and possibilities of autonomy offer avenues to incorporate our ethical system and moral considerations within the decision-making of the 'robot'. Presently, face-recognition is not highly advanced nor faultless, but that will change with time. What will incorporating such technology and coupling it with reconnaissance mean for our military capabilities and our view on its application? Further investigation on such developments is a worthwhile pursuit as it could lead to new thinking and legislation concerning the use of AI under certain circumstances.

As for proliferation, one might say that the rules for export control could be applicable. Highly advanced technological knowledge and systems only have the competitive advantage as long as they are confined to one party. And as long as that party is adhering to IHL and subsequent legislation there should not be any restrictions on the use of these kinds of systems.

Public and political support are essential. Clear communication and well led debate is critical to bring home the message that one should not run away from this difficult task and that support is required to further experimentation and use.

5.4 Line 4: Organization and Leadership

The organization is not yet ready to fully adapt to RAS, the concept of units as they are known by tradition is less relevant and needs rethinking.

Operating with RAS introduces many challenges as described above, with issues needing to be teased out, developed, and implemented. Without strong and visionary leadership and a motivated and forward-thinking organization, this task will be near impossible. In numerous fields, not only new concepts, doctrine, training, and logistics have to be thought through, but also future applications and needs have to be defined. One of the first questions to arise will be the number of RAS needed for certain tasks and military missions. Will RAS be added at a group, platoon or company level? How much RAS will units possess, at platoon and company level, or at HQ level? This will fully depend on its military tasks and what the specific RAS is developed for. However, 'orders of battle' have to be defined and figures have to be given at a certain point in time These kinds of questions will guide developers and producers of RAS and will determine prices (for purchase, lease or using the capabilities).

In order to gain knowledge in this field, the RNLA is already experimenting with different types of RAS. The assigned experimentation unit is free to experiment and gain knowledge on RAS in the broadest sense. This kind of freedom within an organization is a prerequisite to gain essential knowledge on all issues raised. At the same time, a display of leadership attitude is much needed, especially when on the forefront of such revolutionary means such as RAS and AI. At the same time, workshops focused on widening knowledge and diving deeper into such questions as stated here will help designing the (RAS-) units of the future.

When entering the RAS era, developments will accelerate. Once the positive sides of working with and alongside RAS have been proven and showcased, the hunger for more RAS will probably increase. The following questions arise regarding the organization of RAS:

- *How to organize the increasing use and dependency on RAS?*
- *Who has control of this momentum, the technician or the commander?*
- *How to organize distributing tasks and commands to RAS?*
- *How to combine the strong points of RAS and humans in training and in life threatening war-fighting ops?*
- *Will we still conduct 'train as you fight' doctrines?*
- *Introducing RAS requires a change of culture, how does leadership foster this change?*

Here, true leadership must be shown enough reign to allow for experimentation far beyond the 'normal' working arena (again, what is normal when working with RAS?), but, simultaneously, reaping the yield of what has been achieved and transferring it into concrete projects and needs.

Human Talent and Training

Training in military environments serves a purpose. It instills discipline, but mainly it provides proficiency and skills in a benign environment which can then be implemented in volatile, and sometimes life-endangering situations. Training is conducted mostly through a step-up program, beginning with individual military skills, developing to work in small groups, to operations at unit level, to integration within a bigger system, and eventually at the level of a fighting force. With the introduction of RAS, the dimension of training will have to adapt to these systems as well. Questions arise on the set-up of training and how to integrate these new systems with their concepts and doctrines:

- *Will training have to be changed or adapted for RAS to maintain the same level and intensity of 'conventional' training?*
- *How can we fit RAS within the training system? Will it be necessary to train as intensively with RAS as it is without them?*
- *If RAS take up the entire fighting force, do we still need training at the highest and most challenging level for humans?*
- *Will the use of RAS lead to risk avoiding behavior within humans ("why sacrifice myself for a robot?")?*
- *How should human personnel cooperate with RAS, will RAS be considered buddies, threats, assistants, subordinates, or superiors?*
- *And what does that mean for training? How do we train for cooperation with RAS in all above levels of hierarchy? How far do we trust RAS to take up certain tasks?*
- *and vice-versa, do RAS trust the quality of decision making and guidance by humans?*

As this is uncharted territory, the optimal way to find answers on above questions is a careful experimental approach. However, with rapidly developing technology (think of Moore's law), evolutions within RAS and AI might dictate the tempo and leave no room for a 'crawl-walk-run' approach on training. As RAS do not suffer from fatigue and have unending stamina, thinking needs to be done on what this means for teaming up with personnel who have physical limitations. Is it possible for RAS understand these human limitations and take them into account?

When RAS are not used in training, exercises or operations, they must be stored somewhere. Preferably, this storage should be conditioned to prevent RAS from unnecessary exposure to excessive rain, humidity, sun and extreme weather conditions. What are the infrastructural challenges and standards for storing RAS when not deployed/used in training and is specific infrastructure needed for repair and maintenance?

Perhaps not every training with RAS takes place outside. There might be cases when training inside is required or when training exercises take place indoors with the help of Virtual or Augmented Reality techniques. For reasons of Operational Security (OpSec), training must be shielded from outside observers, be it from space, from the air or from the surface of the earth. In such a case, terrain that is closed-off and covered must be available. It follows that the introduction of RAS can easily have infrastructural implications. Further research must be done if special infrastructure must be built and if extra equipment must be acquired.

In order to transport RAS between exercise and operation locations and the places where they are sheltered, they must move or be moved. Some RAS will fit into trucks and can be transported as normal cargo, while other RAS are self-driving or self-flying. We must consider that RAS will frequently use public roads, or move through the air, in/across the water. For all these environments traffic rules apply. That means that RAS will have to be certified for use through the public domain as well. A form of certification will have to take place before RAS can be introduced as military inventory.

In the military, training and proficiency in tasks often lead to increasing competencies, better performances and increasing career chances. This leads to important questions on the implementation of RAS on careers and training:

- *What if essential military manual labor is not conducted by personnel anymore, will they still be able to build their basic skills, or will the set of basic skills completely disappear?*
- *What does this mean for career possibilities?*
- *Do personnel have to be selected for working with RAS, can everybody work with RAS, or does it require extra skills and qualifications?*
- *How do soldiers acquire these skills, will the introduction of RAS lead to the need for a different type of soldier eventually?*

All of these questions will require answers before work with RAS can commence but can only be truly answered whilst working with RAS. Every type of RAS will require its own set of challenges, dilemmas and procedures when they are integrated into working with humans. The man-machine teaming concepts of 'loyal wingmen', 'flocking' and 'swarming' will have to be defined for each type of RAS, mission, and task. It is likely that over time, more definitions and hybrid forms of man-machine teaming will be developed. Depending on certain missions and surrounding circumstances, the concepts can be switched, either temporarily or situationally. Significant flexibility will be required of the soldiers that work with RAS and the operational concepts that come with working them. Special selection criteria probably will have to apply to attract the right kind of qualified personnel.

Logistics and infrastructure

All military equipment must be serviced or repaired occasionally. Independent of the discussion on ownership, the materiel will occasionally be unserviceable, out of order, under repair or under further development. Additionally, the provision of upgrades for either the hardware or the software renders equipment unavailable for certain periods of time. Just as with conventional equipment, there will have to be enough operational systems to continue training and missions alongside planned and unplanned maintenance. Hardware and software maintenance, in principle, do not require specific logistic challenges for a military organization. There are examples that new dimensions are introduced in RAS, one being automatic software upgrades (for example, Tesla and its electric powered cars). How do the military keep check on the software updates, and how much does the updated software still comply to the set ROE? Do RAS still perform in the same way, with the same commands? Have they become 'smarter' by a software update, and what does that mean for procedures, tasks, doctrine and interoperability with other RAS and humans? What further complicates these questions is when RAS are equipped with self-learning software. In this case, updates and improvements are introduced incrementally and almost continually and perhaps even without knowledge of the operator. A system of quality control will have to be implemented within the organization to keep track of enhanced software, improved capabilities, and related (interoperability) issues, like compliance with ROE for example.

Furthermore, within logistical processes, the maintenance interval is often important. The amount of 'flying hours' or operating hours for RAS before depletion will decide the availability of these systems. With the use of big data for logistical support and life-cycle maintenance, the period of time when RAS can be in use can be better scheduled, leading to optimal use of available time and systems. Battle damage repair is another issue to be explored. Depending on the damage, RAS, like any other equipment, is classified as either destroyed or repairable. Moreover, the threshold of damage to RAS that can be sustained before the system should be withdrawn needs to be prescribed. This is critical to the prevention of destruction or capture of potentially sensitive technologies. But to which level of repairable state RAS can be managed in the field or not, must be prescribed. Certain RAS systems can contain either sensitive information or technology or both, which must be salvaged in case of damage, destruction or capture so that it does not fall into the opponents' hands.

RAS require other logistical processes than human-centric solutions, are tasked differently, have different operating cycles, and require different modes of learning. Initially, RAS will likely function as complementary to human operators as part of what will for the time being remain human-centric solutions. Eventually, as RAS mature and armed forces become more familiar with RAS, dedicated RAS-centric solutions will be sought out and found. This will likely imply not only the adaptation of existing human-centric processes and structures (in evolutionary, incremental steps), but also the transformation process, for them to fit into the peculiarities of RAS (in rapid, potentially precarious leaps).

6. Conclusion

Before multiple RAS applications can be effectively applied to the military toolbox, extensive thinking about numerous peculiarities specific to the military must be conducted. In some instances, thinking has to precede the decision making, while in others, decision making cannot be done without practical experiences and experimentation. The former is already taking place, albeit on a small scale. This paper addresses additional viewpoints and questions to carefully consider during the process of acquiring increasing numbers of RAS. At the same time, the questions and the issues raised here are far from complete. The best way forward is to experiment and learn at the same time and to learn simultaneously and continuously.

Based on our assessments, several actions required for the implementation of RAS, were identified and addressed in the lines of development. This could serve as a point of departure when thinking about the implementation of RAS in the armed forces.

The questions explored in this paper will assist in attaining a firmer hold on RAS issues as we venture into uncharted territory. Overall, there is only one way forward in modern warfighting, and that will involve an increased use of RAS and the accompanying AI. It is up to decision makers to make this path as smooth and complete as possible by enhancing the benefits of this technology and addressing the risks and challenges.

Bibliography

Chapter 1

The Military Applicability of Robotic and Autonomous Systems

Airforce Technology. "Mantis MALE Unmanned Aerial Vehicle (UAV)." *Airforce Technology* (blog). Accessed March 1, 2019. https://www.airforce-technology.com/projects/mantis-uav/.

Army Recognition Group. "Ground and Aerial Unmanned Vehicles UGV UAV in the Israeli Army Defence Forces." Army Recognition Group, March 1, 2012. https://www.armyrecognition.com/weapons_defence_industry_military_technology_uk/ground_and_aerial_unmanned_vehicles_ugv_uav_in_the_israeli_army_defence_forces_idf_0103126.html.

———. "Russian Special Forces Have Received Platform-M UGV Unmanned Ground Vehicles." Army Recognition Group, February 1, 2016. https://www.armyrecognition.com/february_2016_global_defense_security_news_industry/russian_special_forces_have_received_platform-m_ugv_unmanned_ground_vehicles_tass_10102163.html.

"Army Warfighting Experiment 2018: Autonomous Warrior." British Army Innovation Technology Book, 2018.

Boulanin, Vincent, and Maaike Verbruggem. "Mapping the Development of Autonomy in Weapon Systems." Stockholm International Peace Research Institute, November 2017.

"Casper 250." Accessed March 1, 2019. http://www.israeli-weapons.com/weapons/aircraft/uav/casper_250/Casper_250.htm.

Defence Industries. "Israeli Firm Revives Old Concept With Advanced Robotics." Defence Industries, October 7, 2016. https://www.defence-industries.com/news/israeli-firm-revives-old-concept.

Feickert, Andrew, Jennifer K. Elsea, Lawrence Kapp, and Laurie A. Harris. *U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI): Considerations for Congress*. Independently published, 2018.

Horowitz, Michael, and Paul Scharre. "Meaningful Human Control in Weapon Systems: A Primer." Center for a New American Security, March 16, 2015. https://www.cnas.org/publications/reports/meaningful-human-control-in-weapon-systems-a-primer.

Hsu, Jeremy. "Real Soldiers Love Their Robot Brethren." *LiveScience*, May 21, 2009. https://www.livescience.com/5432-real-soldiers-love-robot-brethren.html.

Lim, Zhifeng. "The Rise of Robots and the Implications for Military Organizations." Naval Postgraduate School, September 2013. https://apps.dtic.mil/dtic/tr/fulltext/u2/a589729.pdf.

Nathan Fisher, and Gary Gilbert. “Medical Robotic and Autonomous System Technology Enablers for the Multi-Domain Battle 2030-2050.” Small Wars Journal. Accessed March 1, 2019. https://smallwarsjournal.com/jrnl/art/medical-robotic-and-autonomous-system-technology-enablers-for-the-multi-domain-battle-2030-.

National Robotics Engineering Center. “Autonomous Platform Demonstrator.” Carnegie Mellon University. Accessed March 1, 2019. http://www.cmu.edu/nrec/solutions/defense/other-projects/crusher.html.

Reilly, M.B. “Beyond Video Games: New Artificial Intelligence Beats Tactical Experts in Combat Simulation.” University of Cincinnati, June 27, 2016. https://magazine.uc.edu:8443https://magazine.uc.edu/editors_picks/recent_features/alpha.

SAE International. “Automated Driving: Levels of Driving Automation Are Defined in New SAE International Standard J3016.” SAE International, June 18, 2014. https://web.archive.org/web/20170903105244/https://www.sae.org/misc/pdfs/automated_driving.pdf.

Smith, Rich. “The U.S. Navy Spent $744 Million to Build a Robotic Fighter Jet -- and Now Wants to Throw It Away -.” The Motley Fool, May 24, 2015. https://www.fool.com/investing/general/2015/05/24/us-navy-spent-744-million-robotic-fighter-jet.aspx.

UK Ministry of Defence. “Joint Concept Note 1/18: Human-Machine Teaming.” UK Ministry of Defence, May 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/709359/20180517-concepts_uk_human_machine_teaming_jcn_1_18.pdf.

U.S. Army. “The U.S. Army Robotic and Autonomous Systems Strategy.” Maneuver, Aviation, and Soldier Division Army Capabilities Integration Center, March 2017. http://www.arcic.army.mil/app_Documents/RAS_Strategy.pdf.

Wolf, Amelia Mae, and Micah Zenko. “Drones Kill More Civilians Than Pilots Do.” *Foreign Policy*, April 25, 2016. https://foreignpolicy.com/2016/04/25/drones-kill-more-civilians-than-pilots-do/.

Chapter 2

The Ethics of Robotic and Autonomous Systems in a Military Context

"Accountability." In *Business Dictionary*, n.d. http://www.businessdictionary.com/definition/accountability.html.

"AEGIS Weapon System." US Navy Fact File. US Navy, October 1, 2019. https://www.navy.mil/navydata/fact_display.asp?cid=2100&tid=200&ct=2.

"Annex 3-60 – Targeting. Appendix A: Targeting and Legal Consideration. Basic Principles of the Law of War and Their Targeting Implications." U.S. Air Force Doctrine, March 15, 2019. https://www.doctrine.af.mil/Portals/61/documents/Annex_3-60/3-60-D33-Target-LOAC.pdf.

Arkin, Ronald C. *Governing Lethal Behavior in Autonomous Robots*. Boca Raton F.L.: CRC Press Taylor & Francis Group, 2009.

———. "Lethal Autonomous Systems and the Plight of the Non-Combatant." *AISB Quarterly*, July 2013.

Asaro, Peter. "Jus Nascendi, Robotic Weapons and the Martens Clause." In *Robot Law*, by Ryan Calo, A. Michael Froomkin, and Ian Kerr, 2016. https://www.elgaronline.com/view/edcoll/9781783476725/9781783476725.00024.xml.

———. "The Liability Problem for Autonomous Artificial Agents." In *AAAI Spring Symposia*, 2016.

"Australia's System of Control and Applications for Autonomous Weapon Systems." Geneva, 2019. https://www.unog.ch/80256EDD006B8954/(httpAssets)/39A4B669B8AC2111C12583C1005F73CF/$file/CCW_GGE.1_2019_WP.2_final.pdf.

Banta, Benjamin R. "'The Sort of War They Deserve'? The Ethics of Emerging Air Power and the Debate over Warbots." *Journal of Military Ethics* 17, no. 2–3 (July 3, 2018): 156–71. https://doi.org/10.1080/15027570.2018.1551320.

Bendett, S. "In AI, Russia Is Hustling to Catch Up." *Defense One*, April 4, 2018. https://www.defenseone.com/ideas/2018/04/russia-races-forward-ai-development/147178/.

Bhana, Hemant. "By the Book - Good Written Guidance and Procedures Reduce Pilots' Automation Complacency." Aero Safety World, March 2010. https://flightsafety.org/wp-content/uploads/2016/11/asw_mar10_p47-51.pdf.

Bo, Marta, and Taylor Woodcock. "Blog: Lethal Autonomous Weapons, War Crimes, and the Convention on Conventional Weapons." *Asser Institute* (blog), May 28, 2019. https://www.asser.nl/about-the-institute/asser-today/blog-lethal-autonomous-weapons-war-crimes-and-the-convention-on-conventional-weapons/.

Boeing. "Maintenance Program Enhancements." *Boeing*, 2006. http://www.boeing.com/commercial/aeromagazine/articles/qtr_4_06/article_05_2.html.

Boogaard, Jeroen van den. "Proportionality and Autonomous Weapons Systems." Opinio Juris, 2016. http://opiniojuris.org/2016/03/23/proportionality-and-autonomous-weapons-systems/.

Boothby, William, ed. *New Technologies and the Law in War and Peace*. Cambridge University Press, 2018. https://www.cambridge.org/core/books/new-technologies-and-the-law-in-war-and-peace/AA47FC74ABEA568F6971E53EF906601C.

Bouvier, Antoine A. "International Humanitarian Law and the Law of Armed Conflict." Edited by Harvey J. Langholtz. Peace Operations Training Institute, 2012. http://cdn.peaceopstraining.org/course_promos/international_humanitarian_law/international_humanitarian_law_english.pdf.

Boyd, John. "The Essence of Winning and Losing." 1995. https://www.danford.net/boyd/essence.htm.

Boyle, Michael J. "The Legal and Ethical Implications of Drone Warfare." *The International Journal of Human Rights* 19, no. 2 (February 17, 2015): 105–26. https://doi.org/10.1080/13642987.2014.991210.

Boyle v United Techs. Corp. 487 U.S. 500, 510 (1988) (n.d.).

Brenton, Richard, and Eloi Bosse. "The Cognitive Costs and Benefits of Automation." In *RTO-MP-088*, 2002. https://www.researchgate.net/publication/235171183_The_Cognitive_Costs_and_Benefits_of_Automation.

Bryson, Joanna, Mihailis Diamantis, and Thomas Grant. “Of, For, and By the People: The Legal Lacuna of Synthetic Persons.” *Artificial Intelligence and Law* 25, no. 3 (September 30, 2017): 273–91.

Burrell, Jenna. “How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms.” *Big Data & Society* 3, no. 1 (June 1, 2016): 2053951715622512. https://doi.org/10.1177/2053951715622512.

Calloway, Audra. “Army Wargames Shape the Future of Urban Warfare.” U.S.Army, January 3, 2019. https://www.army.mil/article/215731/army_wargames_shape_the_future_of_urban_warfare.

Cassese, Antonio. “The Martens Clause: Half a Loaf or Simply Pie in the Sky?” *European Journal of International Law* 11, no. 1 (2000): 187–216.

Chairperson of the Informal Meeting of Experts. “Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS).” United Nations Office at Geneva, 2016. https://www.unog.ch/80256EDD006B8954/(httpAssets)/DDC13B243BA863E6C1257FDB00380A88/$file/ReportLAWS_2016_AdvancedVersion.pdf.

Charisi, Vicky, Louise Dennis, Michael Fisher, Robert Lieck, Andreas Matthias, Marija Slavkovik, Janina Loh, Alan Winfield, and Roman Yampolskiy. “Towards Moral Autonomous Systems,” November 1, 2017. https://arxiv.org/pdf/1703.04741.pdf.

Chehtman, Alejandro. “New Technologies Symposium: Autonomous Weapons Systems– Why Keeping a ‘Human On the Loop’ Is Not Enough.” *Opinio Juris* (blog), May 8, 2019. https://opiniojuris.org/2019/05/08/new-technologies-symposium-autonomous-weapons-systems-why-keeping-a-human-on-the-loop-is-not-enough/.

Chung, Timothy. “OFFensive Swarm-Enabled Tactics.” Defense Advanced Research Projects Agency (DARPA). Accessed August 14, 2019. https://www.darpa.mil/program/offensive-swarm-enabled-tactics.

Clarke, Arthur C. *Profiles of the Future; an Inquiry into the Limits of the Possible*. New York: Harper & Row, 1973.

Cohen, Stanley. *States of Denial: Knowing about Atrocities and Suffering*. Polity, 2001.

Committee on Legal Affairs. “Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).” European Parliament, May 31, 2016. http://www.europarl.europa.eu/doceo/document/JURI-PR-582443_EN.pdf?redirect.

“Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field. Geneva, 12 August 1949.” International Committee of the Red Cross (ICRC), 1949. https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/ART/365-570004?OpenDocument.

“Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects.” Geneva, 2019. https://www.unog.ch/80256EDD006B8954/(httpAssets)/39A4B669B8AC2111C12583C1005F73CF/$file/CCW_GGE.1_2019_WP.2_final.pdf.

Crawford, Kate, and Meredith Whittaker. “Artificial Intelligence Is Hard to See.” Medium, November 9, 2016. https://medium.com/@katecrawford/artificial-intelligence-is-hard-to-see-a71e74f386db.

Crootof, Rebecca. “War Torts: Accountability for Autonomous Weapons.” *University of Pennsylvania Law Review* 164 (May 2016): 56.

Cummings, Mary. “Automation Bias in Intelligent Time Critical Decision Support Systems.” *American Institute of Aeronautics and Astronautics*, no. Session: IS-15: Human Interaction with Intelligent Systems (September 2004). https://doi.org/10.2514/6.2004-6313.

“Customary IHL - Practice Relating to Rule 14. Proportionality in Attack.” ICRC IHL Database. Accessed June 25, 2019. https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_rul_rule14.

Davison, Neil. “A Legal Perspective :Autonomous Weapon Systems under International Humanitarian Law.” *UNODA Occasional Papers* 30 (2018).

———. “Autonomous Weapon Systems: An Ethical Basis for Human Control?” ICRC Humanitarian Law & Policy Blog, April 3, 2018. https://blogs.icrc.org/law-and-policy/2018/04/03/autonomous-weapon-systems-ethical-basis-human-control/.

Declaration Renouncing the Use, in Time of War, of Explosive Projectiles Under 400 Grammes Weight (1868). https://ihl-databases.icrc.org/ihl/full/declaration1868.

Deeks, Ashley, Noam Lubell, and Daragh Murray. "Machine Learning, Artificial Intelligence, and the Use of Force by States." *Journal of National Security Law & Policy* 10, no. 1 (2019): 1–25.

Development, Concepts and Doctrine Centre. *Human-Machine Teaming*. Joint Concept Note, 1/18. UK Ministry of Defence, 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/709359/20180517-concepts_uk_human_machine_teaming_jcn_1_18.pdf.

———. "Unmanned Aircraft Systems - Joint Doctrine Publication 0-30.2." UK Ministry of Defence, 2017. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/673940/doctrine_uk_uas_jdp_0_30_2.pdf.

Dickinson, Laura. "Contract as a Tool for Regulating Private Military Companies." In *From Mercenaries to Market*, by Simon Chesterman and Chia Lehnardt. Oxford University Press, 2007. https://doi.org/10.1093/acprof:oso/9780199228485.001.0001.

———. *Outsourcing War and Peace*. Yale University Press, 2011. https://yalebooks.yale.edu/book/9780300144864/outsourcing-war-and-peace.

Dickinson, Laura A. "Drones, Automated Weapons, and Private Military Contractors." In *New Technologies for Human Rights Law and Practice*, edited by Molly K. Land and Jay D. Aronson, 93–124. Cambridge University Press, 2018. https://doi.org/10.1017/9781316838952.005.

"Dilbert at War." *The Economist*, June 23, 2014. https://www.economist.com/united-states/2014/06/23/dilbert-at-war.

Docherty, Bonnie. "Banning 'Killer Robots': The Legal Obligations of the Martens Clause." Arms Control Association, October 1, 2018. https://www.armscontrol.org/act/2018-10/features/remarks-banning-%E2%80%98killer-robots%E2%80%99-legal-obligations-martens-clause.

———. "Heed the Call: A Moral and Legal Imperative to Ban Killer Robots." USA: Human Rights Watch, August 21, 2018. https://www.hrw.org/report/2018/08/21/heed-call/moral-and-legal-imperative-ban-killer-robots.

———. "Losing Humanity: The Case against Killer Robots." Human Rights Watch, 2012.

Ekelhof, Merel. "Autonome Wapensystemen: Wat We Moeten Weten over de Toepassing van Het Humanitair Oorlogsrecht En de Menselijke Rol in Militaire Besluitvorming." *Ars Aequi*, March 2018, 193–202.

———. "Autonomous Weapons: Operationalizing Meaningful Human Control." Humanitarian Law & Policy Blog, August 15, 2018. https://blogs.icrc.org/law-and-policy/2018/08/15/autonomous-weapons-operationalizing-meaningful-human-control/.

———. "Lifting the Fog of Targeting: 'Autonomous Weapons' and Human Control through the Lens of Military Targeting." *Naval War College Review* 71, no. 3 (2018): 61–94.

Etzioni, Amitai, and Oren Etzioni. "Pros and Cons of Autonomous Weapons Systems." *Military Review* 97, no. 3 (June 2017): 72–81.

European Commission. "Commission Regulation (EU) No 1321/2014 of 26 November 2014 on the Continuing Airworthiness of Aircraft and Aeronautical Products, Parts and Appliances, and on the Approval of Organisations and Personnel Involved in These Tasks." Office Journal of the European Union, December 17, 2014. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R1321&from=EN.

European Group on Ethics in Science and New Technologies (EGE). *Statement on Artificial Intelligence, Robotics and "autonomous" Systems*. Luxembourg: European Commission Directorate-General for Research and Innovation, 2018.

Evans, Hayley. "Too Early for a Ban: The U.S. and U.K. Positions on Lethal Autonomous Weapons Systems." *Lawfare* (blog), April 13, 2018. https://www.lawfareblog.com/too-early-ban-us-and-uk-positions-lethal-autonomous-weapons-systems.

Feickert, Andrew, Lawrence Kapp, Jennifer K Elsea, and Laurie A Harris. "U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI): Considerations for Congress." U.S. Congressional Research Service, November 20, 2018. https://digital.library.unt.edu/ark:/67531/metadc1442984/.

Floridi, Luciano. "What the Near Future of Artificial Intelligence Could Be." *Philosophy & Technology* 32, no. 1 (March 1, 2019): 1–15. https://doi.org/10.1007/s13347-019-00345-y.

Floridi, Luciano, Jessica Morley, Libby Kinsey, and Anat Elhalal. "From What to How - An Overview of AI Ethics Tools, Methods and Research to Translate Principles into Practices," 2019. https://www.academia.edu/39135750/From_What_to_How-_An_Overview_of_AI_Ethics_Tools_Methods_and_Research_to_Translate_Principles_into_Practices.

Folkman, Joseph. "The '8 Great' Accountability Skills For Business Success." *Forbes*, November 14, 2014. https://www.forbes.com/sites/joefolkman/2014/11/14/how-do-you-score-the-8-great-accountability-skills-for-business-success/#74093a803c11.

Galand, Alexandre, Emilie Hunter, and Ilia Utmelidze. "International Criminal Law Guidelines: Command Responsibility." Case Matrix Network, January 2016. https://www.legal-tools.org/doc/7441a2/pdf/.

Giger, Jean-Christophe, Nuno Piçarra, Patrícia Alves-Oliveira, Raquel Oliveira, and Patricia Arriaga. "Humanization of Robots: Is It Really Such a Good Idea?" *Human Behaviour & Emergining Technologies*, 2019, 111–23.

Gilbert, David. "Russian Weapons Maker Kalashnikov Developing Killer AI Robots." *Vice News*, July 13, 2017. https://news.vice.com/en_us/article/vbzq8y/russian-weapons-maker-kalashnikov-developing-killer-ai-robots.

Gray, H.M., K. Gray, and D.M. Wegner. "Dimensions of Mind Perception." *Science* 315, no. 5812 (2007): 619.

Grishenko, Nikolai. "Российский Подводный Робот Выполнил Боевую Задачу в Сирии." Российская Газета (*RG.RU*), February 22, 2018. https://rg.ru/2018/02/22/rossijskij-podvodnyj-robot-vypolnil-boevuiu-zadachu-v-sirii.html.

Gunkel, David. "Other Things: AI, Robots and Society." In *A Networked Self and Human Augmentics, Artificial Intelligence, Sentience*, 1st ed., 216. Routledge, 2018. https://www.taylorfrancis.com/books/e/9781315202082.

Han, Yi, Benjamin I. P. Rubinstein, Tamas Abraham, Tansu Alpcan, Olivier De Vel, Sarah Erfani, David Hubczenko, Christopher Leckie, and Paul Montague. "Reinforcement Learning for Autonomous Defence in Software-Defined Networking." In *Decision and Game Theory for Security*, edited by Linda Bushnell, Radha Poovendran, and Tamer Başar, 11199:145–65. Springer International Publishing, 2018. https://doi.org/10.1007/978-3-030-01554-1_9.

Hawkins, Andrew. "Everything You Need to Know about the Boeing 737 Max Airplane Crashes." *The Verge*, March 22, 2019. https://www.theverge.com/2019/3/22/18275736/boeing-737-max-plane-crashes-grounded-problems-info-details-explained-reasons.

Henderson-Sellers, Brian, and Julian M. Edwards. "The Object-Oriented Systems Life Cycle." *Communications of the ACM* 33, no. 9 (September 1990): 142–159. https://doi.org/10.1145/83880.84529.

"Hero-400EC Extended-Range Loitering System." Air Force Technology, n.d. https://www.airforce-technology.com/projects/hero-400ec-extended-range-loitering-system/.

Hoek, Freek, Cor van Montfort, and Cees Vermeer. "Enhancing Public Accountability in the Netherlands." *OECD Journal on Budgeting* 5, no. 2 (2005): 70–86.

Hoijtink, Marijn, and Matthias Leese. *Technology and Agency in International Relations*. 1st ed. Routledge, 2019. https://www.taylorfrancis.com/books/e/9780429463143.

Horowitz, Jonathan. "Joint Blog Series: Precautionary Measures in Urban Warfare: A Commander's Obligation to Obtain Information." Humanitarian Law & Policy Blog, January 10, 2019. https://blogs.icrc.org/law-and-policy/2019/01/10/joint-blog-series-precautionary-measures-urban-warfare-commander-s-obligation-obtain-information/.

Horowitz, Michael C., and Paul Scharre. "Meaningful Human Control in Weapon Systems: A Primer." Working paper. Project on Ethical Autonomy. Center for a New American Security (CNAS), March 2015. https://www.files.ethz.ch/isn/189786/Ethical_Autonomy_Working_Paper_031315.pdf.

Hsu, Jeremy. "Real Soldiers Love Their Robot Brethren." *Live Science*, May 21, 2009. https://www.livescience.com/5432-real-soldiers-love-robot-brethren.html.

Hughes, Joshua. "No, Autonomous Weapon Systems Are Not Unlawful under the Martens Clause." Medium, August 21, 2018. https://medium.com/@jghughes1991/no-autonomous-weapon-systems-are-not-unlawful-under-the-martens-clause-2653d18790e9.

International Committee of the Red Cross ICRC. "Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control?" Geneva, Switzerland: International Committee of the Red Cross (ICRC), 2018. https://www.icrc.org/en/download/file/69961/icrc_ethics_and_autonomous_weapon_systems_report_3_april_2018.pdf.

International Law Commission. "Responsibility of States for Internationally Wrongful Acts." United Nations, 2001. http://legal.un.org/ilc/texts/instruments/english/draft_articles/9_6_2001.pdf.

JASON, and The MITRE Corporation. "Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to DoD," January 2017. https://fas.org/irp/agency/dod/jason/ai-dod.pdf.

Johansson, Linda. "Ethical Aspects of Military Maritime and Aerial Autonomous Systems." *Journal of Military Ethics* 17, no. 2–3 (July 3, 2018): 140–55. https://doi.org/10.1080/15027570.2018.1552512.

Jones, Bruce, Charles T Call, Daniel Toubolets, and Jason Fritz. "Managing the New Threat Landscape: Adapting the Tools of International Peace and Security." *Brookings Institute*, Foreign Policy at Brookings, September 2018, 23.

Kania, Elsa. "China's Strategic Ambiguity and Shifting Approach to Lethal Autonomous Weapons Systems." Lawfare, April 17, 2018. https://www.lawfareblog.com/chinas-strategic-ambiguity-and-shifting-approach-lethal-autonomous-weapons-systems.

———. "China's Strategic Ambiguity and Shifting Approach to Lethal Autonomous Weapons Systems." Lawfare, April 17, 2018. https://www.lawfareblog.com/chinas-strategic-ambiguity-and-shifting-approach-lethal-autonomous-weapons-systems.

———. "The Critical Human Element in the Machine Age of Warfare." Foundatio. *Courier* (blog), n.d. https://www.stanleyfoundation.org/articles.cfm?id=867&title=A-Happy-Place--to-Be-a-Cow.

Kania, Elsa B. "Battlefield Singularity: Artificial Intelligence, Military Revolution, and China's Future Military Power." Washington, D.C.: Center for a New American Security (CNAS), November 2017. https://s3.amazonaws.com/files.cnas.org/documents/Battlefield-Singularity-November-2017.pdf?mtime=20171129235805.

"Killer Robots - Learn." Campaign to Stop Killer Robots. Accessed March 7, 2019. https://www.stopkillerrobots.org/learn/.

"Killer Robots and the Concept of Meaningful Human Control." Human Rights Watch, 2016. https://www.hrw.org/sites/default/files/supporting_resources/robots_meaningful_human_control_final.pdf.

"Killer Robots Fail Key Moral, Legal Test: Principles and Public Conscience Call for Preemptive Ban." *Human Rights Watch* (blog), August 21, 2018. https://www.hrw.org/news/2018/08/21/killer-robots-fail-key-moral-legal-test.

Klare, Michael T. "Autonomous Weapons Systems and the Laws of War." Arms Control Association, March 2019. https://www.armscontrol.org/act/2019-03/features/autonomous-weapons-systems-laws-war.

Koohi v. United States, 976 F.2d 1328, 1336–37 (9th Cir. 1992) (n.d.).

Krishman, Armin. *Killer Robots: Legality and Ethicality of Autonomous Weapons*. Routledge, 2016. https://books.google.nl/books?id=ZS0HDAAAQBAJ&pg=PA140&lpg=PA140&dq=military+Emotional+attachment+robots&source=bl&ots=PNKx4VSYIl&sig=ACfU3U2ewDF0-bAWt3Aty3oOb-d96co0xg&-hl=fr&sa=X&ved=2ahUKEwjR27nkqo7hAhVBJ1AKHatzCdU4FBDoATAFegQIBBAB#v=onepage&q=military%20Emotional%20attachment%20robots&f=false.

Kumar, Padma, Sharafat Hussain, Alexis Espiritu, Lj Marzo, and Rosa Rakavono. "Algorithms, Flowcharts, Data Types and Pseudocode." Accessed August 14, 2019. https://www.academia.edu/34581869/2._ALGORITHMS_FLOWCHARTS_DATA_TYPES_AND_PSEUDOCODE_2.1_ALGORITHMS.

Legality of the Threat or Use of Nuclear Weapons (Advisory Opinion of 8 July 1996), 1996 Reports 226 (ICJ 1996).

Lehman, M. M. "Programs, Life Cycles, and Laws of Software Evolution." *Proceedings of the IEEE* 68, no. 9 (September 1980): 1060–76. https://doi.org/10.1109/PROC.1980.11805.

Leveringhaus, Alex. "Autonomous Weapons Mini-Series: Distance, Weapons Technology and Humanity in Armed Conflict." Humanitarian Law & Policy Blog, October 6, 2017. https://blogs.icrc.org/law-and-policy/2017/10/06/distance-weapons-technology-and-humanity-in-armed-conflict/.

Lewis, Larry. "AI and Autonomy in War: Understanding and Mitigating Risks." CNA Center for Autonomy and AI, n.d. https://www.cna.org/CNA_files/PDF/Understanding-Risks.pdf.

Lijn, Jaïr van der, and Stefanie Ros. “Peacekeeping Contributor Profile: The Netherlands.” Providing for Peacekeeping, January 2014. http://www.providingforpeacekeeping.org/2014/04/08/contributor-profile-the-netherlands/.

Lin, Patrick, George Bekey, and Keith Abney. “Autonomous Military Robotics: Risk, Ethics, and Design.” Ethics + Emerging Sciences Group at California Polytechnic State University, December 20, 2008. https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1001&context=phil_fac.

———. “Robots in War: Issues of Risk and Ethics.” In *Ethics and Robotics*, edited by R. Capurro and M. Nagenborg, 49–67. AKA Verlag Heidelberg, 2009. https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1010&context=phil_fac.

Liu, Hin-Yan. “Contract Law as Cover: Curtailing the Scope of Private Military and Security Contractor Responsibilities.” *The Ashgate Research Companion to Outsourcing Security: The Role of the Market in 21st Century Warfare (Joakim Berndtsson and Christopher Kinsey Eds.)*, 2016. https://www.academia.edu/15301427/Contract_Law_as_Cover_Curtailing_the_Scope_of_Private_Military_and_Security_Contractor_Responsibilities.

Long, Drake. “China Releases Video of 56-Boat Drone Swarm near Hong Kong.” *TheDefensePost*, February 6, 2018. https://thedefensepost.com/2018/06/02/china-56-boat-drone-swarm-hong-kong/.

Lu, Denise, Allison McCann, Jin Wu, and K. K. Rebecca Lai. “From 8,600 Flights to Zero: Grounding the Boeing 737 Max 8.” *The New York Times*, March 13, 2019. https://www.nytimes.com/interactive/2019/03/11/world/boeing-737-max-which-airlines.html.

Malik, Swati. “Autonomous Weapon Systems: The Possibility and Probability of Accountability.” *Wisconsin International Law Journal* 35, no. 3 (n.d.): 34.

Marchant, Gary, Ronald C. Arkin, Braden Allenby, Edward T. Barrett, Jason Borenstein, Lyn Gaudet, Orde Kittrie, et al. “International Governance of Autonomous Military Robots.” *Columbia Science & Technology Law Review* 272 (2011).

Marra, William C., and Sonia K. Mcneil. “Understanding ‘The Loop’: Regulating the Next Generation of War Machines.” *Harvard Journal of Law & Public Policy* 36, no. 3 (May 2013): 1139–85.

Matthias, Andreas. “The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata.” *Ethics and Information Technology* 6, no. 3 (September 2004): 175–83.

McIntyre, Alison. “Doctrine of Double Effect.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2019. Metaphysics Research Lab, Stanford University, 2019. https://plato.stanford.edu/archives/spr2019/entries/double-effect/.

McLean, Wayne. “Drones Are Cheap, Soldiers Are Not: A Cost-Benefit Analysis of War.” The Conversation. Accessed June 28, 2019. http://theconversation.com/drones-are-cheap-soldiers-are-not-a-cost-benefit-analysis-of-war-27924.

Melzer, Nils. “Keeping the Balance between Military Necessity and Humanity – a Response to Four Critiques of the ICRC’s Interpretive Guidance on the Notion of Direct Participation in Hostilities.” *New York University Journal of International Law and Politics* 42 (2010): 831–916.

Miller, Christopher A., and Raja Parasuraman. “Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control.” *Human Factors* 49, no. 1 (February 1, 2007): 57–75. https://doi.org/10.1518/001872007779598037.

“Milrem Robotics Delivered Two THeMIS UGVs to the Dutch Army.” Private. Milrem Robotics, May 28, 2019. https://milremrobotics.com/milrem-robotics-delivered-two-themis-ugvs-to-the-dutch-army/.

Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. “The Ethics of Algorithms: Mapping the Debate.” *Big Data & Society* 3, no. 2 (December 1, 2016): 2053951716679679. https://doi.org/10.1177/2053951716679679.

Mohanty, Bedavyasa. “Lethal Autonomous Dragon: China’s Approach to Artificial Intelligence Weapons.” *ORF* (blog), November 15, 2017. https://www.orfonline.org/expert-speak/lethal-autonomous-weapons-dragon-china-approach-artificial-intelligence/.

Mucic et al (“Celebici”), No. IT-96-21-T (ICTY November 16, 1998).

Musgrave, Shawn. "Inside 'Liberty City,' Homeland Security's Site for Testing Urban Drones." Vice Motherboard, June 19, 2015. https://www.vice.com/en_us/article/wnj9nq/inside-liberty-city-homeland-securitys-site-for-testing-urban-drones.

Netherlands Advisory Council on International Affairs. "Autonomous Weapon Systems: The Need for Meaningful Human Control." Netherlands Advisory Council on International Affairs, October 2015. https://aiv-advies.nl/8gr#government-responses.

Parasuraman, Raja, and Dietrich Manzey. "Complacency and Bias in Human Use of Automation: An Attentional Integration." *Human Factors* 52, no. 3 (October 2010): 381–410. https://doi.org/10.1177/0018720810376055.

Pastor, E, C Barrado, P Royo, J Lopez, and E Santamaria. "An Open Architecture for the Integration of UAV Civil Applications, Aerial Vehicles." In *Aerial Vehicles*, 511–36. IntechOpen, 2009. https://www.intechopen.com/books/aerial_vehicles/an_open_architecture_for_the_integration_of_uav_civil_applications.

Preece, Alun. "Asking 'Why' in AI: Explainability of Intelligent Systems – Perspectives and Challenges." *Intelligent Systems in Accounting, Finance and Management* 25, no. 2 (April 19, 2018). https://doi.org/10.1002/isaf.1422.

"Preparing for More Urban Warfare." *The Economist*, January 25, 2018. https://www.economist.com/special-report/2018/01/25/preparing-for-more-urban-warfare.

Press, Michael. "Of Robots and Rules: Autonomous Weapon Systems in the Law of Armed Conflict." *Georgetown Journal of International Law* 48 (2017): 1337–66.

Prosecutor v Kavishema, No. ICTR-95-1-T (ICTR May 21, 1999).

Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) (1977).

"Reframing Autonomous Weapons Systems." In *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems (A/IS)*. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, n.d. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_reframing_autonomous_weapons_v2.pdf.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," February 16, 2016. http://arxiv.org/abs/1602.04938.

Robert, Lionel. "The Growing Problem of Humanizing Robots." *International Robotics & Automation Journal* 3, no. 1 (2017). https://medcraveonline.com/IRATJ/IRATJ-03-00043.pdf.

Roff, Heather M., and David Danks. "'Trust but Verify': The Difficulty of Trusting Autonomous Weapons Systems." *Journal of Military Ethics* 17, no. 1 (January 2, 2018): 2–20. https://doi.org/10.1080/15027570.2018.1481907.

"RQ-11 Raven Unmanned Aerial Vehicle." Army Technology, n.d. https://www.army-technology.com/projects/rq11-raven/.

Sandoz, Yves, Christophe Swinarski, and Bruno Zimmermann. "Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949." International Commission of the Red Cross, 1987. https://perma.cc/5XKM-QQYV.

Santoni de Sio, Filippo, and Jeroen van den Hoven. "Meaningful Human Control over Autonomous Systems: A Philosophical Account." *Frontiers in Robotics and AI* 5, no. 15 (2018). https://doi.org/10.3389/frobt.2018.00015.

Scharre, Paul. *Army of None: Autonomous Weapons and the Future of War*. WW Norton & Co, 2018.

Scharre, Paul, and Michael Horowitz. "Meaningful Human Control in Weapon Systems: A Primer." Center for New American Security, March 2015. https://www.files.ethz.ch/isn/189786/Ethical_Autonomy_Working_Paper_031315.pdf.

Schmitt, Michael, and Jeffrey Thurnher. "'Out of the Loop': Autonomous Weapon Systems and the Law of Armed Conflict." *Harvard National Security Journal* 4, no. 231 (2013). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2212188.

Schwarz, Elke. "Intelligent Weapons Systems and Meaningful Human Control: An Uneasy Alliance," 2019. Working Paper - available from author.

———. “The (Im)Possibility of Meaningful Human Control for Lethal Autonomous Weapon Systems.” International Committee of the Red Cross. *Humanitarian Law & Policy* (blog), August 29, 2018. https://blogs.icrc.org/law-and-policy/2018/08/29/im-possibility-meaningful-human-control-lethal-autonomous-weapon-systems/.

Scott, Andrew, Jose Solorzano, Jonathan Moyer, and Barry Hughes. “Modeling Artificial Intelligence and Exploring Its Impact.” Frederick S. Pardee Center for International Futures Josef Korbel School of International Studies University of Denver, May 2017. https://pardee.du.edu/sites/default/files/ArtificialIntelligenceIntegratedPaper_V6_clean.pdf.

Sharkey, Noel. “Guidelines for the Human Control of Weapons Systems.” ICRAC, April 2018. https://www.icrac.net/icrac-working-paper-3-ccw-gge-april-2018-guidelines-for-the-human-control-of-weapons-systems/.

———. “Killer Robots From Russia Without Love.” *Forbes*, November 28, 2018. https://www.forbes.com/sites/noelsharkey/2018/11/28/killer-robots-from-russia-without-love/.

Sharkey, Noel E. “The Evitability of Autonomous Robot Warfare.” *International Review of the Red Cross*, Comments and Opinions, 94, no. 886 (Summer 2012): 787–99. https://doi.org/10.1017/S1816383112000732.

Shilo, Liron. “Speaking of Responsibility: Autonomous Weapon Systems, State and Individual Responsibility.” Georgetown Law, Institute for Technology, Law & Policy. *Georgetown Tech* (blog), n.d. https://www.georgetowntech.org/blogfulltext/2017/5/1.

Sparrow, Rob. “Ethics as a Source of Law: The Martens Clause and Autonomous Weapons.” Humanitarian Law & Policy Blog, November 14, 2017. https://blogs.icrc.org/law-and-policy/2017/11/14/ethics-source-law-martens-clause-autonomous-weapons/.

Sparrow, Robert. “Building a Better Warbot: Ethical Issues in the Design of Unmanned Systems for Military Applications.” *Science and Engineering Ethics* 15, no. 2 (2009): 169–187. https://doi.org/10.1007/s11948-008-9107-0.

Spiegeleire, Stephan De, Matthijs Maas, and Tim Sweijs. *Artificial Intelligence and the Future of Defense*. The Hague, The Netherlands: The Hague Centre For Strategic Studies, 2017. https://hcss.nl/sites/default/files/files/reports/Artificial%20Intelligence%20and%20the%20Future%20of%20Defense.pdf.

“Statement of the Head of the Russian Federation Delegation, Director of the Department for Nonproliferation and Arms Control of the Russian Ministry for Foreign Affairs V.Yermakov at the Meeting of the State-Parties of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons on Item 7 of the Agenda ‘General Exchange of Views’, Geneva, November 21, 2018.” The Ministry of Foreign Affairs of the Russian Federation, November 22, 2018. http://www.mid.ru/main_en/-/asset_publisher/G51iJnfMMNKX/content/id/3415655.

Stewart, Emily. “The Boeing 737 Max 8 Crashes and Controversy, Explained.” *Vox*, March 13, 2019. https://www.vox.com/2019/3/12/18262359/boeing-737-max-controversy-faa-trump.

Strawser, Bradley Jay. “Moral Predators: The Duty to Employ Uninhabited Aerial Vehicles.” *Journal of Military Ethics* 9, no. 4 (December 1, 2010): 342–68. https://doi.org/10.1080/15027570.2010.536403.

Suchman, Lucy. *Human-Machine Reconfigurations: Plans and Situated Actions*. 2nd ed. Cambridge University Press, 2006. https://www.cambridge.org/core/books/humanmachine-reconfigurations/9D53E602BA9BB5209271460F92D00EFE.

“The Montreux Document - On Pertinent International Legal Obligations and Good Practices for States Related to Operations of Private Military and Security Companies during Armed Conflict.” *International Committee of the Red Cross*, 2009, 48.

“The Position Paper Submitted by the Chinese Delegation to CCW 5th Review Conference.” Geneva: United Nations Office at Geneva, n.d. https://www.unog.ch/80256EDD006B8954/(httpAssets)/DD1551E60648CEBBC125808A005954FA/$file/China%27s+Position+Paper.pdf.

Torossian, Bianca, Frank Bekkers, Tim Sweijs, Michel Roelen, Alen Hristov, and Salma Atalla. “Paper on the Military Applicability of Robotic and Autonomous Systems.” The Hague Centre for Strategic Studies, Paper forthcoming - available from authors 2019.

Travis, Gregory. “How the Boeing 737 Max Disaster Looks to a Software Developer.” *IEEE Spectrum*, April 18, 2019. https://spectrum.ieee.org/aerospace/aviation/how-the-boeing-737-max-disaster-looks-to-a-software-developer.

United Nations, Department of Economic and Social Affairs, Population Division. “World Urbanization Prospects: The 2018 Revision,” 2018.

United Nations Institute for Disarmament Research (UNIDIR. “The Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence.” United Nations, 2018. http://www.unidir.ch/files/publications/pdfs/the-weaponization-of-increasingly-autonomous-technologies-artificial-intelligence-en-700.pdf.

———. “The Weaponization of Increasingly Autonomous Technologies: Considering How Meaningful Human Control Might Move the Discussion Forward.” United Nations, 2014. http://www.unidir.ch/files/publications/pdfs/considering-how-meaningful-human-control-might-move-the-discussion-forward-en-615.pdf.

“Unmanned Systems Integrated Roadmap FY 2011-2036.” US Department of Defense, 2011. https://fas.org/irp/program/collect/usroadmap2011.pdf.

US Department of Defense. “Directive 3009.09.” US Government, November 21, 2012. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf.

Verdiesen, Ilse. “Agency Perception and Moral Values Related to Autonomous Weapons: An Empirical Study Using the Value-Sensitive Design Approach. Masters Thesis.” TU Delft, 2017.

Vincent, James. “Giving Robots ‘Personhood’ Is Actually about Making Corporations Accountable.” The Verge, January 19, 2017. https://www.theverge.com/2017/1/19/14322334/robot-electronic-persons-eu-report-liability-civil-suits.

Weisgerber, Marcus. “What’s in the House NDAA?; Pentagon’s 3D-Mapping Service; New Marine One, Weed Whacker; and More.” Defense One, May 10, 2018. https://www.defenseone.com/business/2018/05/global-business-brief-may-10-2018/148116/.

Worcester, Maxim. “Autonomous Warfare – A Revolution in Military Affairs.” *ISPSW Strategy Series: Focus on Defense and International Security*, no. 340 (April 2015): 6.

Wu, Xiang-Hu, Ming-Cheng Qu, Zhi-Qiang Liu, and Jian-Zhong Li. “Research and Application of Code Automatic Generation Algorithm Based on Structured Flowchart.” *Journal of Software Engineering and Applications* 4 (2011): 534–45. https://doi.org/10.4236/jsea.2011.49062.

Chapter 3

Managing RAS: The Need for New Norms and Arms Control

ECP: Platform voor de InformatieSamenleving. “Aanpak Begeleidingsethiek in Het NRC,” December 5, 2019. https://ecp.nl/actueel/aanpak-begeleidingsethiek-in-het-nrc/.

Adler, Emanuel, and Vincent Pouliot. “International Practices.” *International Theory* 3, no. 1 (February 18, 2011): 1–36.

“Artificial Intelligence and Life in 2030.” Stanford University, September 2016. https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai100report10032016fnl_singles.pdf.

Barrie et al., Douglas. “2019. Capturing Technology. Rethinking Arms Control. Conference Reader.” German Federal Foreign Office, March 15, 2019. https://rethinkingarmscontrol.de/wp-content/uploads/2019/03/2019.-Capturing-Technology.Rethinking-Arms-Control_-Conference-Reader.pdf.

Blok, Stef. “Betreft Motie Koopmans c.s. over Beheersing van de Productie, Plaatsing, Verspreiding En Inzet van Nieuwe Potentiële Massavernietigingswapens.” Ministerie van Buitenlandse Zaken, September 20, 2019. https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2019/09/20/kamerbrief-over-nieuwe-potentiele-massavernietigingswapens/kamerbrief-over-nieuwe-potentiele-massavernietigingswapens.pdf.

———. “Speech by Stef Blok, Minister of Foreign Affairs, at Amsterdam Drone Week, 6 December 2019.” presented at the Amsterdam Drone Week, Amsterdam, December 6, 2019. https://www.government.nl/documents/speeches/2019/12/06/speech-by-minister-stef-blok-at-amsterdam-drone-week.

Braithwaite, John, and Peter Drahos. *Global Business Regulation*. Cambridge: Cambridge University Press, 2000.

Chavannes, Esther, and Amit Arkhipov-Goyal. “Towards Responsible Autonomy.” The Ethics of Robotic and Autonomous Systems in a Military Context. The Hague: The Hague Centre for Strategic Studies, September 2019. https://hcss.nl/sites/default/files/files/reports/Towards%20Responsible%20Autonomy%20-%20The%20Ethics%20of%20RAS%20in%20a%20Military%20Context.pdf.

Chavannes, Esther, Klaudia Klonowska, and Tim Sweijs. “Governing Autonomous Weapon Systems: Expanding the Solution Space, from Scoping to Applying.” The Hague: The Hague Centre for Strategic Studies, February 2020.

Davis Cross, Mai’a K. “Re-Thinking Epistemic Communities Twenty Years Later.” *Review of International Studies* 38, no. 1 (January 2013): 137–60.

Davis, Zachary. “Artificial Intelligence on the Battlefield – Implications for Deterrence and Surprise.” *PRISM: The Journal of Complex Operations* 8, no. 2 (2019).

Dekker, Brigitte, and Maaike Okano-Heijmans. “Emerging Technologies and Competition in the Fourth Industrial Revolution: The Need for New Approaches to Export Control.” *Strategic Trade Review* 6, no. 9 (Winter/Spring 2020): 53–67.

———. “The US–China Trade–Tech Stand-Off.” The Hague: The Clingendael Institute, August 2019.

Delcker, Janosch, and Andrew Gray. “Top UN Official: It’s Not Too Late to Curb AI-Powered Weapons.” *POLITICO*, February 13, 2020. https://www.politico.eu/article/top-un-official-its-not-too-late-to-curb-ai-powered-weapons/.

“Draft Report of the 2019 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems.” Geneva: United Nations GGE, August 21, 2019. https://www.unog.ch/80256EDD006B8954/(httpAssets)/5497DF9B01E5D9CFC125845E00308E44/$file/CCW_GGE.1_2019_CRP.1_Rev2.pdf.

Evans, Hayley, and Natalie Salmanowitz. “Lethal Autonomous Weapons Systems: Recent Developments.” *Lawfare* (blog), March 7, 2019. https://www.lawfareblog.com/lethal-autonomous-weapons-systems-recent-developments.

Ezell, Stephen, and Caleb Foote. “How Stringent Export Controls on Emerging Technologies Would Harm the U.S. Economy.” Washington, DC: Information Technology & Innovation Foundation, May 2019.

Ford, Christopher Ashley. “Coalitions of Caution: Building a Global Coalition Against Chinese Technology-Transfer Threats.” presented at the FBI-Department of Commerce Conference on Counter-Intelligence and Export Control, Indianapolis, Indiana, September 13, 2018.

Haas, Peter M. “Introduction: Epistemic Communities and International Policy Coordination.” *International Organizations* 46, no. 1 (Winter 1992): 1–35.

European Commission. “High-Level Expert Group on Artificial Intelligence,” October 4, 2019. https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence.

Kallenborn, Zachary, and Philipp C. Bleek. “Drones of Mass Destruction: Drone Swarms and the Future of Nuclear, Chemical, and Biological Weapons.” *War on the Rocks* (blog), February 14, 2019. https://warontherocks.com/2019/02/drones-of-mass-destruction-drone-swarms-and-the-future-of-nuclear-chemical-and-biological-weapons/.

Kania, Elsa. “China’s Strategic Ambiguity and Shifting Approach to Lethal Autonomous Weapons Systems.” *Lawfare* (blog), April 17, 2018. https://www.lawfareblog.com/chinas-strategic-ambiguity-and-shifting-approach-lethal-autonomous-weapons-systems.

Keck, Margaret E., and Kathryn Sikkink. “Transnational Advocacy Networks in International and Regional Politics.” *International Social Science Journal* 51, no. 159 (March 1999): 89–101.

Kelly, Kevin. “The Three Breakthroughs That Have Finally Unleashed AI on the World.” *Wired*, October 27, 2014. https://www.wired.com/2014/10/future-of-artificial-intelligence/.

Lewis, Larry. “Killer Robots Reconsidered: Could AI Weapons Actually Cut Collateral Damage?” *Bulletin of the Atomic Scientists* (blog), January 10, 2020. https://thebulletin.org/2020/01/killer-robots-reconsidered-could-ai-weapons-actually-cut-collateral-damage/.

Leys, Nathan. “Autonomous Weapon Systems and International Crises.” *Strategic Studies Quarterly*, no. Spring 2018 (2018): 51.

MacCarthy, Mark. “AI Needs More Regulation, Not Less.” *Brookings* (blog), March 9, 2020. https://www.brookings.edu/research/ai-needs-more-regulation-not-less/.

Masuhr, Niklas. “AI in Military Enabling Applications.” Edited by Fabien Merz. *CSS Analyses in Security Policy*, no. 251 (October 2019): 1–4.

Netherlands Advisory Council on International Affairs. “Autonomous Weapon Systems: The Need for Meaningful Human Control.” Netherlands Advisory Council on International Affairs, October 2015. https://aiv-advies.nl/8gr#government-responses.

The Campaign To Stop Killer Robots. “New European Poll Shows Public Favour Banning Killer Robots,” November 13, 2019. https://www.stopkillerrobots.org/2019/11/new-european-poll-shows-73-favour-banning-killer-robots/.

North Atlantic Treaty Organization. “Arms Control, Disarmament and Non-Proliferation in NATO.” North Atlantic Treaty Organization, November 28, 2019. http://www.nato.int/cps/en/natohq/topics_48895.htm.

Pauwelyn, Joost, Ramses A. Wessel, and Jan Wouters. “The Stagnation of International Law.” Working paper. Leuven: Leuven Centre for Global Governance Studies, 2012. https://research.utwente.nl/en/publications/the-stagnation-of-international-law.

Renic, Neil C. “Death of Efforts to Regulate Autonomous Weapons Has Been Greatly Exaggerated.” *Bulletin of the Atomic Scientists*, December 18, 2019. https://thebulletin.org/2019/12/death-of-efforts-to-regulate-autonomous-weapons-has-been-greatly-exaggerated/.

Tesner, Sandrine, and Georg Kell. *The United Nations and Business: A Partnership Recovered*. New York: St. Martin’s Press, 2000.

The Campaign to Stop Killer Robots. “The Campaign To Stop Killer Robots,” 2018. https://www.stopkillerrobots.org/.

The United Nations. “2019 Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS).” United Nations Geneva, n.d. https://www.unog.ch/80256EE600585943/(httpPages)/5535B644C2AE8F28C1258433002BBF14.

Torossian, Bianca, Frank Bekkers, Tim Sweijs, Michel Roelen, Alen Hristov, and Salma Atalla. "The Military Applicability of Robotic and Autonomous Systems." Security. HCSS Security. The Hague: The Hague Centre For Strategic Studies (HCSS), March 1, 2019.

Tsingou, Eleni. "Transnational Policy Communities and Financial Governance: The Role of Private Actors in Derivatives Regulation." Working paper. Coventry: Centre for the Study of Globalisation and Regionalisation, January 2003.

United Nations Group of Governmental Experts. "Draft Report of the 2019 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems." Geneva: United Nations GGE, August 21, 2019. https://www.unog.ch/80256EDD006B8954/(httpAssets)/5497DF9B01E5D9CFC125845E00308E44/$file/CCW_GGE.1_2019_CRP.1_Rev2.pdf.

Utting, Peter. "Codes in Context: TNC Regulation in an Era of Dialogues and Partnerships." Briefing. The Corner House, February 2002.

———. "UN-Business Partnerships: Whose Agenda Counts?," 1–18. Oslo, Norway: United Nations Research Institute for Social Development, 2000.

Verbruggen, Maaike. "The Role of Civilian Innovation in the Development of Lethal Autonomous Weapon Systems." *Global Policy* 10, no. 3 (September 2019): 338–42.

Woolf, Amy F., Mary Beth D. Nikitin, and Paul K. Kerr. "Arms Control and Nonproliferation: A Catalog of Treaties and Agreements." Congressional Research Service, March 18, 2019. https://crsreports.congress.gov/product/pdf/RL/RL33865.

Chapter 4

Effective Stakeholder Cooperation during the lifecycle of Robotic and Autonomous Systems

Allen, Greg, and Taniel Chan. “Artificial Intelligence and National Security.” Belfer Center for Science and International Affairs, July 2017. https://www.belfercenter.org/sites/default/files/files/publication/AI%20NatSec%20-%20final.pdf.

“Army Cyber Training and Education within Finabel Member States.” FINABEL European Army Interoperability Center, 2019. http://finabel.org/wp-content/uploads/2019/01/FQ_Cyber_Training_and_Education_Web2.pdf.

Bromley, Mark, and Giovanni Maletta. “The Challenge of Software and Technology Transfers to Non-Proliferation Efforts,” n.d.

Chavannes, Esther, and Amit Arkhipov-Goyal. “Towards Responsible Autonomy: The Ethics of Robotic and Autonomous Systems in a Military Context.” The Hague: The Hague Centre for Strategic Studies, September 2019. https://www.hcss.nl/sites/default/files/files/reports/Towards%20Responsible%20Autonomy%20-%20The%20Ethics%20of%20RAS%20in%20a%20Military%20Context.pdf.

Chavannes, Esther, Klaudia Klonowska, and Tim Sweijs. “Governing Autonomous Weapon Systems.” Den Haag: The Hague Centre for Strategic Studies, February 13, 2020.

Clarke, Duncan L. “The Arrow Missile: The United States, Israel and Strategic Cooperation.” *Middle East Journal* 48, no. 3 (1994). https://www.jstor.org/stable/4328717.

Kortenkamp, David, and Reid Simmons. “Robotic Systems Architectures and Programming.” In *Springer Handbook of Robotics*, edited by Bruno Siciliano and Oussama Khatib, 187–206. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. https://doi.org/10.1007/978-3-540-30301-5_9.

Perlo-Freeman, Sam, and Elisabeth Sköns. “The Private Military Services Industry.” Stockholm International Peace Research Institute, September 2008. https://www.sipri.org/publications/2008/sipri-insights-peace-and-security/private-military-services-industry.

Platts, Jon, Mary Cummings, and Rory Kerr. “Applicability of STANAG 4586 to Future Unmanned Aerial Vehicles.” In *AIAA Infotech@Aerospace 2007 Conference and Exhibit*. Rohnert Park, California: American Institute of Aeronautics and Astronautics, 2007. https://doi.org/10.2514/6.2007-2753.

“Robotic and Autonomous Systems of Systems Architecture.” Army Science Board. The Army Science Board, Department of the Army, January 15, 2017. https://apps.dtic.mil/dtic/tr/fulltext/u2/1058366.pdf.

“Safety, Unintentional Risk and Accidents in the Weaponization of Increasingly Autonomous Technologies.” The United Nations Institute for Disarmament Research, 2016. https://www.unidir.org/files/publications/pdfs/safety-unintentional-risk-and-accidents-en-668.pdf.

Slijper, Frank, Alice Beck, and Daan Kayser. “State of AI: Artificial Intelligence, the Military and Increasingly Autonomous Weapons.” Pax for Peace, April 2019.

“The U.S. Army: Robotic and Autonomous Systems Strategy.” U.S. Army Training and Doctrine Command, March 2017. https://www.tradoc.army.mil/Portals/14/Documents/RAS_Strategy.pdf.

Verbruggen, Maaike. “The Role of Civilian Innovation in the Development of Lethal Autonomous Weapon Systems.” *Global Policy* 10, no. 3 (September 2019): 338–42. https://doi.org/10.1111/1758-5899.12663.

Wasiak, Radek. “What Is the Incentive in Insurance Premiums?” European Agency for Safety and Health at Work, November 16, 2009. https://osha.europa.eu/sites/default/files/seminars/documents/Radek%20Wasiak%20insurance%20premiums.pdf.

The Hague Centre for Strategic Studies

info@hcss.nl

hcss.nl

Address:
Lange Voorhout 1
2514EA
The Hague
The Netherlands