Clingendael
Netherlands Institute of International Relations

The Hague Centre for Strategic Studies

HCSS Security

Managing RAS: The Need for New Norms and Arms Control

Hugo Klijn and Maaike Okano-Heijmans (Clingendael Institute)

With contributions from Bianca Torossian (HCSS)

“A key lesson of history is that effective regulation focuses on prioritizing the effects of the new technology, rather than the ephemeral technology itself”
-Tom Wheeler, visiting fellow Brookings Institution

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.

Contents

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 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

[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.

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.

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.

[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.

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 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

[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.

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.

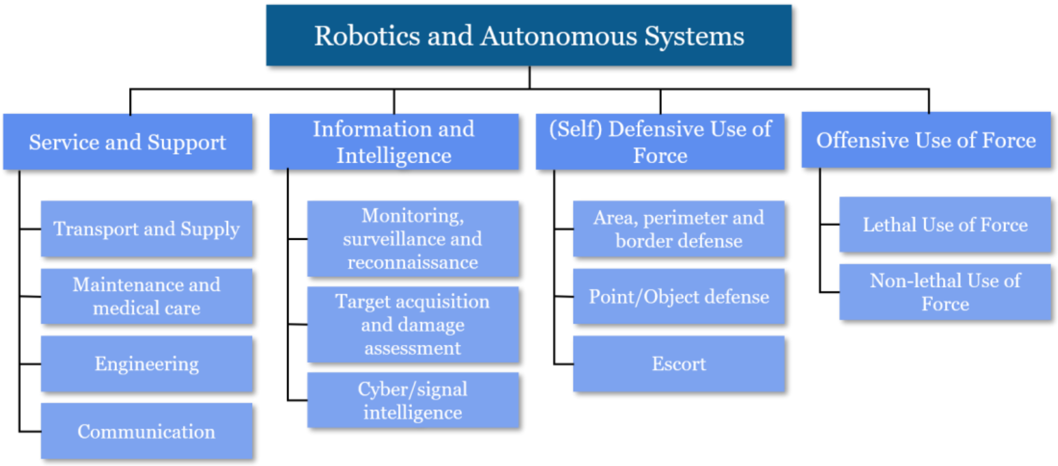

Robotics and Autonomous Systems
Service and Support
Transport and Supply
Maintenance and medical care
Engineering
Communication
Information and Intelligence
Monitoring, surveillance and reconnaissance
Target acquisition and damage assessment
Cyber/signal intelligence
(Self) Defensive Use of Force
Area,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

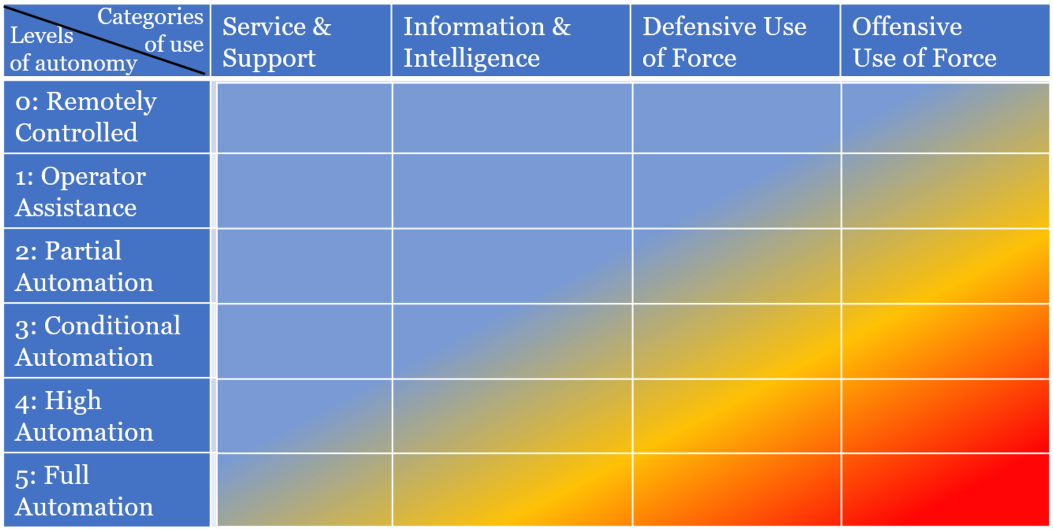
Figure 1: Categorization of RAS.[6]

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

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

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]

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:

[8] Authors' compilation.

[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."

> *"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 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

[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.

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 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

[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/.

[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').

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.

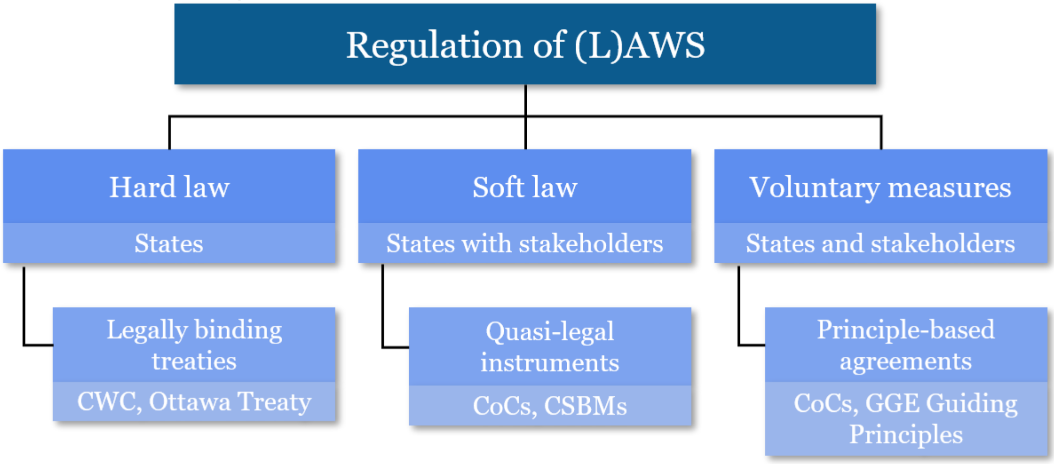


Figure 3: The three mechanisms for regulating (L)AWS.[18]

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

[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.

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]

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

[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.

[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/.

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 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

[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/.
[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).

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.

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

[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.

(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 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

[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.
[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).

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.

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.

Table 1: SWOT analysis of approaches to regulating (L)AWS.[35]

	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[36] • 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

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 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'.

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

[36] 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.

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

[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.

states at risk,[45] 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.

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

[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

[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).

of protecting US national security and 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

[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.

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 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

[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.

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.

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.

Bibliography

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.