UNIDIR

The Weaponization of Increasingly Autonomous Technologies:

Autonomous Weapon Systems and Cyber Operations

No. 7

UNIDIR RESOURCES

Acknowledgements

Support from UNIDIR's core funders provides the foundation for all of the Institute's activities. In addition, dedicated funding for the activities leading to this paper was received from the Governments of Canada, Germany, Ireland, the Netherlands and Switzerland.

UNIDIR would like to acknowledge the thoughtful contributions of the participants in a November 2015 meeting on cyber, AI and increasingly autonomous technologies convened by UNIDIR: David Atkinson, John Borrie, Aude Fleurant, Adam Henschke, Sean Legassick, Patrick Lin, Ryder McKeown, Nils Melzer, Richard Moyes, Jean-Marc Rickli, Heather Roff, Eneken Tikk-Ringas, and Kerstin Vignard. Particular thanks are extended to Patrick Lin for his substantive input, meeting moderation, and synthesis. UNIDIR would also like to acknowledge the contributions of those experts and interviewees who have requested to remain unnamed. This report was drafted by Kerstin Vignard.

About the Project "The Weaponization of Increasingly Autonomous Technologies"

Given that governments have a responsibility to create or affirm sound policies about which uses of autonomy in weapon systems are legitimate—and that advances in relevant technologies are also creating pressure to do so—UNIDIR's work in this area is focused on what is important for States to consider when establishing policy relating to the weaponization of increasingly autonomous technologies. See http://bit.ly/UNIDIR_Autonomy for Observation Papers, audio files from public events, and other materials.

This is the seventh in a series of UNIDIR papers on the weaponization of increasingly autonomous technologies. UNIDIR has purposefully chosen to use the word "technologies" in order to encompass the broadest relevant categorization. In this paper, this categorization includes machines (inclusive of robots and weapons) and systems of machines (such as weapon systems), as well as the knowledge practices for designing, organizing and operating them.

About UNIDIR

The United Nations Institute for Disarmament Research—an autonomous institute within the United Nations—conducts research on disarmament and security. UNIDIR is based in Geneva, Switzerland, the centre for bilateral and multilateral disarmament and non-proliferation negotiations, and home of the Conference on Disarmament. The Institute explores current issues pertaining to the variety of existing and future armaments, as well as global diplomacy and local tensions and conflicts. Working with researchers, diplomats, government officials, NGOs and other institutions since 1980, UNIDIR acts as a bridge between the research community and governments. UNIDIR's activities are funded by contributions from governments and foundations.

Note

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The views expressed in this publication are the sole responsibility of UNIDIR. They do not necessarily reflect the views or opinions of the United Nations or UNIDIR's sponsors.

www.unidir.org

Contents

Acronyms and abbreviations

AEG	Automatic Exploit Generation
AI	Artificial Intelligence
AWS	Autonomous Weapons Systems
CCW	Convention on Certain Conventional Weapons
DARPA	Defense Advanced Research Projects Agency
DoD	Department of Defense (United States)
DoT&E	Department of Defense's Operational Test & Evaluation Directorate (United States)
EMP	Electromagnetic Pulse
GGE	Group of Governmental Experts
ICRC	International Committee of the Red Cross
ICT	Information and Communication Technology
IHL	International Humanitarian Law
LAWS	Lethal Autonomous Weapons Systems
R&D	Research and Development

I. International discussions on autonomous weapon systems and cyber operations

The international discussions on autonomous weapons systems (AWS) focus on conventional weapon systems. Other technologies have not been widely present in the discussion thus far.[1] However, military interest in autonomy is not limited to purely conventional systems—autonomy is relevant for intangible cyber operations[2] as well.

Autonomy-enhancing technological innovations in both physical and digital systems are advancing at a rapid pace. Despite the clear relevance of autonomy for both areas, the international discussions on these issues are held in different multilateral forums with virtually no overlap between the participating experts and policy practitioners. While both subjects are being discussed in formats known as Groups of Governmental Experts (GGEs), the modalities and mode of operation of these groups are completely different.

Starting in 2014, lethal autonomous weapon systems (LAWS) have been taken up as an issue in an arms control framework by the High Contracting Parties of the Convention on Certain Conventional Weapons (CCW). Since then, annual week-long "informal meetings of experts" have discussed and debated various concerns, positions and potential policy responses. In late 2016, the CCW established a formal GGE, which was set to meet in 2017 with a mandate to "explore and agree on possible recommendations on options related to emerging technologies in the area of LAWS", and consider "identification of characteristics and elaboration of a working definition of LAWS".[3] They also noted that further consideration should be given to the "risks posed by cyber operations in relation to LAWS".

Within the CCW framework, International Humanitarian Law (IHL) has been at the forefront of the discussions—not only concerning the legality of LAWS, but also on issues of responsibility and accountability for the use of these weapons (with weapon reviews in the context of Article 36 of the Additional Protocol I to the Geneva Conventions receiving particular attention).[4] There has also been a strong emphasis on emerging norms. Some have highlighted the necessity of human control or judgment when considering the development, deployment and use of LAWS. In this regard, the

[1] Some governments have weighed the commonality between autonomous weapon systems and other autonomous technologies. For instance, the United States Department of Defense Directive 3000.09 on "Autonomy in Weapons Systems" specifically stipulates that the Directive does not apply to "autonomous or semi-autonomous cyber systems for cyberspace operations". This implies an acknowledgement that, intuitively at least, these varying systems with greater degrees of autonomy may be considered together. According to people familiar with the development of the Directive, it was a pragmatic decision to exclude cyber, not a principled one about classification or category boundaries. Accounting for the special issues that arise in cyber would have delayed that directive, which was already urgently needed to clarify policy on emerging robotic systems. United States Department of Defense, "Autonomy in Weapons Systems", Department of Defense Directive 3000.09, section 2.b, 21 November 2012. Available from http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf.

[2] As there is not yet standard terminology in this field, this paper uses the term "cyber operations" rather than "cyber weapons", "cyber arms", "cyber bombs" or other descriptions.

[3] Draft recommendations by the Informal Meetings of Experts are available from http://www.reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2016/meeting-experts-laws/documents/DraftRecommendations_15April_final.pdf; the Final Document of the Fifth Review Conference of the CCW is available from http://www.reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2016/RevCon/documents/final-document.pdf. For an overview of the main concerns, characteristics and definitional approaches, see UNIDIR, "The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches: a Primer", UNIDIR Resources no. 6, 2017. Available from http://bit.ly/UNIDIR_AWSPrimer.

[4] See https://ihl-databases.icrc.org/ihl/WebART/470-750045?OpenDocument.

introduction of the term "meaningful human control" has served as a point of significant discussion.[5]

Building on the 2013 report issued by the UN Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns,[6] civil society engagement has helped to drive this issue to the top of the international agenda and put pressure on policy-makers to take action. Members of the scientific and technical communities, as well as some private sector actors, have taken a proactive interest in establishing clear norms and possibly also legally binding standards.[7] The CCW discussions have thus far included considerable non-governmental expertise, with both invited experts and civil society actors participating in the meetings.

In contrast, the international security implications of cyber operations (known in the UN as "Developments in the Field of Information and Telecommunications in the Context of International Security"[8]) have been on the UN agenda for far longer than AWS. The first Russian-sponsored UN First Committee resolution about "information security" was tabled in 1998.[9]

The discussions on cyber security under the auspices of the United Nations have occurred in closed door, limited membership GGEs established at the request of the First Committee. The first GGE on cyber was held in 2004–2005, and it failed to reach consensus. However, the second GGE (2009–2010) was able to deliver a substantive report, followed by two further successful GGEs in 2012–2013 and 2014–2015. The fifth GGE worked from August 2016 to June 2017, but it ultimately failed to reach a consensus outcome.

The cyber GGEs have confirmed the applicability of international law to cyberspace, encouraged a variety of confidence- and capacity-building measures, and recommended several norms to describe responsible State behaviour in this domain.

Unlike the GGE established by the CCW, the GGEs on cyber security have been relatively small (15–25 States).[10] The cyber GGEs are closed, with no observers (not even other governments, nor relevant international or regional organizations, industry or non-governmental experts are allowed to attend). In addition, concepts like meaningful human control or discussions on Article 36 obligations on the testing of the means and methods of warfare have been absent from the cyber GGEs' deliberations. More broadly, there has been no significant civil society engagement on the international security dimensions of cyber operations.[11]

---

[5] See for instance Article 36, *Key elements of meaningful human control*, Article 36, 2016. Available from http://www.article36.org/publications/#kr; Heather Roff, *Meaningful Human Control, Artificial Intelligence and Autonomous Weapons,* Article 36, 2016. Available from http://www.article36.org/publications/#kr; and UNIDIR, *The Weaponization of Increasingly Autonomous Technologies: Considering How Meaningful Human Control Might Move the Discussion Forward*, UNIDIR Resources, no. 2, 2014. Available from http://unidir.org/files/publications/pdfs/considering-how-meaningful-human-control-might-move-the-discussion-forward-en-615.pdf.

[6] See Cristof Heyns, *Report of the Special Rapporteur on extrajudicial summary or arbitrary executions*, United Nations Human Rights Council, A/HRC/23/47, 2013. Available from www.ohchr.org/Documents/HRBodies/HRCouncil/ RegularSession/Session23/A-HRC-23-47_en.pdf.

[7] See for instance Future of Life Institute, "Autonomous Weapons: An Open Letter from AI & Robotics Researchers", 28 July 2015. Available from http://futureoflife.org/open-letter-autonomous-weapons/; and Future of Life Institute "An Open Letter to The United Nations Convention On Certain Conventional Weapons", 2017. Available from https://futureoflife.org/autonomous-weapons-open-letter-2017/.

[8] See United Nations, Office for Disarmament Affairs, "Developments in the Field of Information and Telecommunications in the Context of International Security", *Fact Sheet*, 2015. Available from https://unoda-web.s3-accelerate.amazonaws.com/wp-content/uploads/2015/07/Information-Security-Fact-Sheet-July2015.pdf.

[9] United Nations, "Developments in the field of information and telecommunications in the context of international security", United Nations Document A/RES/53/70, 1999. Available from http://undocs.org/A/RES/53/70.

[10] For more information about the composition and working methods of the cyber GGEs, see UNIDIR and CSIS, *Report of the International Security Cyber Issues Workshop Series*, 2016, pp. 4–7. Available from http://www.unidir.org/files/publications/pdfs/report-of-the-international-security-cyber-issues-workshop-series-en-656.pdf.

[11] Other cyber-related issues, such as privacy and surveillance, continue to be the focus of considerable attention by civil society groups.

In sum, these two important international discussions have much in common, notably concerning the need to explore appropriate legal and normative frameworks to apply to technological developments. In both discussions, definitions have proven elusive, and existing legal and normative concepts are challenged. There are technical commonalities as well, particularly the intangibility of cyber operations, and at the end of the day it is computer code that will make conventional weapon systems increasingly autonomous.

Despite the commonalities, these international policy discussions have nearly no interaction or overlap. They are followed by different policy practitioners, usually from different sections within Ministries of Foreign Affairs. They are held in different forums, with different levels of transparency. Discussions are for the most part limited to the single technology, and the different GGEs have tended to describe these technologically driven domains as if the technology in question were evolving in isolation. The result is that considerations on potentially significant interactions may be neglected.

Box 1. Examples of increasing autonomy in cyber systems

Increasing autonomy in both offensive and defensive cyber operations is attractive for many of the same reasons as for conventional operations: harnessing ever-greater speed of response, predictive abilities, decision support, and the identification and exploitation of adversaries' vulnerabilities. One example of increasing autonomy in cyber operations is the technique known as Automatic Exploit Generation (AEG), the purpose of which is to "automatically find bugs and generate working exploits".[1]

The 2016 Defense Advanced Research Projects Agency (DARPA) Grand Cyber Challenge was explicitly dedicated to increasing autonomy in cyber operations: "The need for automated, scalable, machine-speed vulnerability detection and patching is large and growing fast as more and more systems—from household appliances to major military platforms—get connected to and become dependent upon the internet. … Machines were challenged to find and patch within seconds—not the usual months—flawed code that was vulnerable to being hacked, and find their opponents' weaknesses before the defending system did."[2] In essence, the challenge was to automate the process of identifying vulnerabilities, simultaneously patching one's own while exploiting the vulnerabilities of other systems.

This well-publicized competition builds upon considerable speculation in the media about government interest in development of autonomous cyber operations.[3] Edward Snowden, for example, alleged that the United States National Security Agency's MonsterMind programme was on its way to becoming an autonomous cyber system:

> *But there were indications it could also include an automated strike-back capability, allowing it to instantly initiate a counterstrike at a piece of malware's source. An error in such an autonomous system, Snowden pointed out, could lead to an accidental war. "What happens when the algorithms get it wrong? ... We're opening the doors to people launching missiles and dropping bombs by taking the human out of the decision chain."*[4]

A standard component of network security includes monitoring threats, recognizing patterns of attack and even anticipating them. Increasingly sophisticated algorithms both detect and respond to potential network threats in the ICT (information and communication technology) environment.[5] Already, due to the processing power and their speed, these "automated" responses are outside real-time human observation or control. At what point would an "automated" response to a cyber threat become an autonomous response?

Notes

1. An "exploit" is "a flaw in hardware or software that is vulnerable to hacking or other cyberattacks", or "a piece of software that takes advantage of such a flaw to compromise a computer system or network." See http://www.dictionary.com/browse/exploit. For a history of AEG research and a fuller explanation of AEG, see Thanassis Avgerinos, et al., "Automatic Exploit Generation" in *Communications of the ACM*, vol. 57, no. 2, 2014, pp. 74–84. Available from https://users.ece.cmu.edu/~dbrumley/pdf/Avgerinos%20et%20al._2014_Automatic%20Exploit%20Generation.pdf.

2. See http://www.darpa.mil/news-events/2016-08-04.

3. Kim Zetter, "Meet Monstermind, The NSA Bot That Could Wage Cyberwar Autonomously", *Wired*, 13 August 2014. Available from http://www.wired.com/2014/08/nsa-monstermind-cyberwarfare/; and Heather Roff, "Monstermind or the Doomsday Machine? Autonomous Cyberwarfare", *Duck of Minerva*, 13 August 2014. Available from http://duckofminerva.com/2014/08/monstermind-or-the-doomsday-machine-autonomous-cyberwarfare.html.

4. James Bamford, "What @Snowden Told Me About the NSA's Cyberweapons", *Foreign Policy*, 29 Sept 2015. Available from http://foreignpolicy.com/2015/09/29/what-snowden-told-me-about-the-nsa-offensive-capabilities/. Read the transcript in its entirety at James Bamford and Tim De Chant, "Edward Snowden on Cyber Warfare", *NOVA Next*, 8 January 2015. Available from http://www.pbs.org/wgbh/nova/next/military/snowden-transcript/.

5. For an example of how artificial intelligence is already deployed in cyber defence in the civilian sector, see Alfred Ng, "Stop cyberattacks. Just add robots", *CNET*, 1 September 2017. Available from https://www.cnet.com/news/cyberattacks-artificial-intelligence-ai-hackers-defcon-black-hat/.

II. Overlap between the two domains

Considering the linkages between increasing autonomy in physical systems and in cyber operations reflects an evolution in our understanding about increasing autonomy in weapon systems, away from purely hardware and mechanical elements towards more software/code-dependent systems. In order to help refine the CCW discussion on LAWS—and to ensure that policy responses are both adequate and sound—it is first worth considering common elements in the two domains.[12]

Dependence on complex learning algorithms and artificial intelligence

Greater autonomy in decision making requires not only greater processing power and increasingly advanced components, such as better sensors, but also more capable software. It requires more sophisticated cognitive capabilities to observe and respond to external stimuli appropriately, as well as the ability to create ever more sophisticated models of the world.

> ***Autonomous intelligent agents*** *can be purely software, or integrated into a physical system ('robots')—the difference lies mainly in the environment in which the agent operates: while purely software agents live in what we call 'cyberspace', robots can sense and interact with the same physical environment that we live in. … [T]he similarities between software agents and robots are relevant, given that even in a robot the embedded software … is at the heart of its behavior and capabilities.*[13]

Machine learning, a technique of artificial intelligence (AI), is what will distinguish automated weapons from those with significant (decision-making) autonomous features. Advances in machine learning are progressing at a pace that was unimaginable just a few years ago. In 2016, DeepMind's AlphaGo beat a human champion at Go, having learned to play by analysing human matches. A mere 18 months later, AlphaGo Zero used the technique of reinforcement learning to *teach itself* how to play Go, and in a matter of days outperformed AlphaGo.[14] According to DeepMind, the latest techniques in the field of reinforcement learning made AlphaGo Zero "more powerful than previous versions of AlphaGo as it is no longer constrained by the limits of human knowledge." With innovation occurring at such a pace, governments are both challenged and pressured to consider how to respond to the potential weaponization of autonomous intelligent agents—whether physical or virtual.

Dominance of the private sector

The private sector is the dominant developer of both cyber technologies and increasing autonomy through applications of AI. The private sector has the human capital, conducts the most research and development (R&D), owns much of the relevant infrastructure, and brings to the market products used around the world for civilian purposes.[15] Traditionally, peaceful applications of

[12] Some early analysis on the intersection on cyber and autonomous weapons includes Caitriona Heinl, "National Security Implication of increasingly Autonomous Technologies: Defining Autonomy, and military and cyber-related implications" (parts 1 and 2), RSIS (2015); and Kenneth Anderson, "Comparing the Strategic and Legal Features of Cyberwar, Drone Warfare, and Autonomous Weapon Systems", *The Briefing*, 27 February 2015.

[13] Alessandro Guarino, "Autonomous Intelligent Agents in Cyber Offence", in K. Podins, J. Stinissen, M. Maybaum (eds.), *5th International Conference on Cyber Conflict*, NATO CCD COE Publications, 2013.

[14] See https://deepmind.com/blog/alphago-zero-learning-scratch/.

[15] While estimates vary, investment in artificial intelligence by the private sector far outpaces governmental investment. For example, a recent report by McKinsey estimated that "tech giants spent $20 billion to $30 billion on AI in 2016, with 90 percent of this spent on R&D and deployment, … . Machine learning, as an enabling technology, received the largest share of both internal and external investment." A few governments have announced significant commitments to public sector support for AI development: for example, in July 2017 the Chinese government announced its "Next Generation Artificial Intelligence Development Plan",

defence technologies were “spun off” to the civilian sector. Today, many private sector advanced tech developments are “spun on” to the defence sector. And although some of the uses of these technologies have significant international security implications, the international security discussions on cyber and autonomous weapons remain dominated by States, with no formal role for the private sector.

In both cyber and AI, the private sector has taken a very public lead in normative development around the uses of these technologies. In 2016, Microsoft, for example, suggested a set of norms for responsible behaviour in cyberspace for both governments and industry,[16] and more recently proposed a “Digital Geneva Convention”.[17] The purpose of these suggested norms was “to advance trust in the global ICT ecosystem through development of ‘rules of the road’ for nation-states engaged in cyber operations, as well as industry actors impacted by these activities.” In the field of AI, industry initiatives such as the Partnership on AI (an industry coalition comprising Amazon, Apple, DeepMind, Facebook, Google, IBM and Microsoft), as well as not-for-profits such as Open AI, AI Now and the Future of Life Institute, are leading the establishment of industry- and research norms concerning AI safety and the use of AI for the benefit of humanity. In both cyber and AI, a considerable number of industry leaders and technical experts are publicly urging governmental restraint on the “weaponization” of these technological developments, such as through an open letter to the CCW issued by a group of researchers and experts in 2017.[18]

Interest, capacity and experience of malicious actors

Sophisticated AWS are typically characterized as so technologically advanced that non-state actors or criminals will not have access to them—although concerns have been raised about non-state actors using much cruder AWS, based on commercially available materials. The cyber domain is already utilized by malicious actors for a variety of motivations—ideological, economic, or simply because they have the ability to do so. There is a robust and growing market for cyber exploits—software designed to attack digital vulnerabilities—and thus economic incentive for malicious actors to discover, use or sell these new vulnerabilities.

A challenge to traditional arms control approaches

What does proliferation mean when a single copy of software can be replicated countless times, by anyone, nearly instantaneously, at minimal cost and be transferred anywhere in the world in less than seconds? Whether open source code for learning algorithms, or exploits for sale on the dark web, these tools are widely accessible to States and non-state actors alike.

Historically the arms control community has attempted to control the spread and harmful use of dual-use/sensitive materials through conventions, such as the Chemical Weapons Convention and the Biological Weapons Convention, as well as through initiatives whereby groups of States control

committing to be the “premier global AI innovation center” by 2030; in October the United Arab Emirates established the world’s first State Minister for AI. How these initiatives translate into financial investment remains to be seen. See McKinsey Global Institute, *Artificial Intelligence The Next Digital Frontier?,* 2017. Available from https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx.

[16] Scott Charney et al., “From Articulation to Implementation: Enabling progress on cybersecurity norms”, Microsoft, 2016. Available from https://mscorpmedia.azureedge.net/mscorpmedia/2016/06/Microsoft-Cybersecurity-Norms_vFinal.pdf.

[17] Brad Smith, “The Need for a Digital Geneva Convention”, Transcript of Keynote Address at the RSA Conference 2017. Available from https://mscorpmedia.azureedge.net/mscorpmedia/2017/03/Transcript-of-Brad-Smiths-Keynote-Address-at-the-RSA-Conference-2017.pdf.

[18] Future of Life Institute (2017), *op. cit.*

or restrict access to particular items or materials due to proliferation concerns. Examples include the Missile Technology Control Regime, the Nuclear Suppliers Group and the Wassenaar Arrangement.

Controlling malicious or harmful applications of software—such as malware, intrusion software, or "weaponized" code (for example Stuxnet)—does not lend itself to these traditional approaches. Control lists would be outdated before they could even be agreed upon. Moreover, if these exploits were to target vulnerabilities previously unknown to their developers and operators, known as "zero-day exploits",[19] it would be impossible to control in advance because they are by fiat unknown until triggered, and worthless once announced, discovered or used. Finally, as seen in 2017, the "stockpiling" of cyber exploits by governments presents a particularly high risk, and current methods for cyber "stockpile management" are clearly insufficient.[20]

As mentioned above, for the most part it is the private sector, not governments, that controls much of the materials, conducts the R&D and brings to market the majority of these dual-use technologies. They represent the leading sectors in many developed economies. Some governments have already articulated their concern that regulation of dual-use software and AI applications would result in being denied access to these technologies, locked out of important high-tech sectors, or inhibit R&D for civilian applications of increasing autonomy. Together these factors make traditional responses, such as export control regimes, even less likely to succeed.

A closer examination of unsuccessful efforts to control or verify dual-use materials might yield useful insights, such as the failed attempt to negotiate a mechanism to verify the Biological Weapons Convention, or that of the Wassenaar Arrangement's attempt to limit the spread of intrusion software, which had the unintended consequence of hobbling legitimate R&D on anti-surveillance and vulnerability discovery in the cyber realm.

Difficulties for testing and verification

Increasing autonomy in both the conventional and cyber realm challenge existing ways to assess the legality of new means and methods of warfare, as required by Article 36 of the Additional Protocol I to the Geneva Conventions.[21] This is further complicated in the case of learning systems where learning is not "frozen" at deployment, as these systems will evolve as they interact with their environment.

Verification in the cyber domain poses both technical and strategic challenges. How could any sort of regulation be verified since determining whether code is "militarized" or not might be near impossible? On the strategic level, it is difficult to imagine that States or corporations would be willing to permit inspection of their code or algorithms as a verification measure.

With systems based on artificial neural networks,[22] there is no way at present to test, verify and validate that a system will perform as requested. Artificial neural networks take in vast amounts of data inputs, and feed these through a series of layers, with each node in the network feeding into another node in another layer. Once the system produces an output, it is too convoluted and complex to trace back and determine why the system decided what it did. Issues surrounding

[19] For more on zero-day exploits, see for instance https://www.fireeye.com/current-threats/what-is-a-zero-day-exploit.html.
[20] The stockpiling of cyber exploits by governments received considerable attention in May 2017 when a cyber exploit called "EternalBlue", developed by a government agency, was stolen and leaked online by a hacking group known as the Shadow Brokers and combined with the ransomware known as WannaCry.
[21] See, for example, Vincent Boulanin, *Implementing Article 36 Weapon Reviews in the Light of Increasing Autonomy in Weapon Systems*, SIPRI, 2015. Available from https://www.sipri.org/sites/default/files/files/insight/SIPRIInsight1501.pdf.
[22] Artificial neural networks can be understood as "computing systems inspired by the biological neural networks that constitute animal brains." For more information, see https://en.wikipedia.org/wiki/Artificial_neural_network.

transparency, interpretability, validation and some form of verification for such systems are among the most important research questions in the AI community today.

As the Director of the United States Department of Defense's Operational Test and Evaluation (DoT&E) Directorate recently acknowledged: "As our systems become even more complex, *and autonomous*, continuous and integrated testing will be necessary"[23] [emphasis added]. These techniques, however, are far from perfected.

In both AI applications and in cyber, capabilities—of States and non-state actors alike—are hard to gauge, and therefore to verify. Where stability measures such as verification once involved counting the number of physical weapons in a stockpile, cyber and AI capabilities are not as amenable to quantitative assessment. Increasing autonomy in conventional weapon systems, in military decision-making aids, and in cyber operations will require new approaches to transparency and verification—perhaps using techniques like distributed ledger technologies[24] (such as Blockchain)—if they are to be regulated in a relevant way.

[23] United States Department of Defense, Operational Test and Evaluation Directorate, *FY 2016 Annual Report,* p. xvii. Available from http://www.dote.osd.mil/pub/reports/FY2016/pdf/other/2016DOTEAnnualReport.pdf.

[24] Distributed ledger technology can be understood as "a consensus of replicated, shared, and synchronized digital data geographically spread across multiple sites, countries, or institutions." For more information, see https://en.wikipedia.org/wiki/Distributed_ledger.

III. What sort of interplay is there between increasing autonomy in conventional systems and cyber operations?

While there are numerous ways in which increasing autonomy may interact between conventional and virtual systems, three are highlighted here as particularly relevant to the CCW's consideration of LAWS:

- Cyber operations as a driver for increasing autonomy in conventional weapon systems as well as a countermeasure;
- Unintended interactions and emergent behaviours between increasingly autonomous systems; and
- The vulnerability of increasingly autonomous weapon systems to cyber operations.

Cyber operations as a driver and as a countermeasure

Advances and interest in increasingly autonomous conventional weapon systems can drive advances and interest in cyber operations, as the latter would be an effective countermeasure to these physical systems. Unlike a kinetic counter-strike, a cyber operation could be a zero-day exploit resident in the system, lying in wait and triggered with no advance warning. While it would be expected that due to this risk, increasingly autonomous conventional systems will be "hardened systems",[25] all systems have exploitable vulnerabilities. Increasingly hardened systems will inevitably drive development of increasingly sophisticated cyber operations—both for defensive and offensive purposes, thus replicating the traditional escalatory drive of development of new measures and countermeasures. This could be destabilizing, given that in these two strategically important areas there are not yet shared understandings of norms for responsible state behaviour, of how international law applies, nor any international regulation in place that could slow the pace of an arms race.

Unintended interactions and emergent behaviours

While some experts have raised concerns about the risks of unintended interactions and emergent behaviours in AWS,[26] few have considered the risks of unintended interactions between cyber operations and AWS.

The speed at which an autonomous system operates can create new risks. As with the occasional stock market "flash crashes", different algorithms—and even systems with very little autonomy—may interact in unforeseen ways before a human has time to intervene. "Society's techno-social systems are becoming ever faster and more computer-orientated. However, far from simply generating faster versions of existing behaviour, … this speed-up can generate a new behavioural regime as humans lose the ability to intervene in real time."[27] The lead author of a 2013 study of

[25] In computing, "hardening is usually the process of securing a system by reducing its surface of vulnerability, which is larger when a system performs more function". For more information, see https://en.wikipedia.org/wiki/Hardening_(computing).
[26] For a detailed examination of unintentional risk in increasingly autonomous systems, see UNIDIR, *Safety, Unintentional Risk and Accidents in the Weaponization of Increasingly Autonomous Technologies*, UNIDIR Resources no. 5, 2016. Available from http://www.unidir.org/files/publications/pdfs/safety-unintentional-risk-and-accidents-en-668.pdf; and Paul Scharre, *Autonomous Weapons and Operational Risk*, 2016, Center for a New American Security. Available from https://s3.amazonaws.com/files.cnas.org/documents/CNAS_Autonomous-weapons-operational-risk.pdf?mtime=20160906080515.
[27] Neil Johnson et al., "Abrupt rise of new machine ecology beyond human response time", Nature Scientific Reports 3, Article number: 2627 (2013). Available from https://www.nature.com/articles/srep02627.

flash crashes in financial markets described this as a "machine ecology beyond human response time" and stated that "we need to better understand the collective behavior of these interacting systems if we're to avoid problems like microcrashes."[28]

Flash crashes are not errors *per se*: these transactions are simply the result of two (or more) algorithms following their own "correct" logic. Emergent effects (unplanned and unintended) arise from interaction between the systems, and these effects are by definition unpredictable, so our ability to plan for how to mitigate their consequences is poor.

Other cases exist in which algorithms try to outcompete each other, often with absurd results, such as a "flash spike" in pricing.[29] The oft-cited example is of two book sellers on a well-known online marketplace, both using an automatic pricing algorithm, which resulted in a developmental biology textbook, which normally retails for less than USD 40, to be priced at over USD 23 million before someone noticed and interrupted the cycle.[30] Clearly, much more is at stake in military systems. An unexpected auto-escalation by competing AI systems (whether embedded in physical weapon systems or their supporting ICT infrastructure, such as decision aids), could result in a "flash conflict", analogous to other flash events in automated and autonomous systems. While such a flash could occur unintentionally, it could also be done deliberately, such as in the form of a cyber operation designed to trigger this type of event.

A flash incident could even be the "friendly fire" sort. Within a military's arsenal, different systems will have different suppliers, different algorithms optimized for different things, and these systems may exhibit unforeseen behaviour when they interact with one another. Testing interoperability and compatibility will be technically challenging, costly and time consuming. At the same time, having the same code across too many systems means that any vulnerability in the code is a vulnerability replicated in each and every weapon. This sets up a balance to be struck between a diversity of approaches to minimize systemic risks, and a uniformity in standards to better ensure interoperability and integration.

The vulnerability of increasingly autonomous weapon systems to cyber operations

New technology always has potential for new vulnerabilities, or for known vulnerabilities to be exploited in novel ways. Everything that runs on software is vulnerable to attack or manipulation.

Identifying potential risks posed by cyber operations, as well as use-cases in autonomous systems, is complicated because the systems themselves are complex, even more so in "systems of systems".[31] Beyond existing systems, anticipating use-cases and vulnerabilities in *future* systems is even more challenging, as it involves predictions about quickly evolving technologies, coupled with the challenge of anticipating the unpredictable effects of emergent systems. However, it is worth considering the vulnerabilities and risks that we already have seen in existing weapon systems as

[28] George Dvorsky, "A new digital ecology is evolving, and humans are being left behind", 11 September 2013. Available from http://io9.gizmodo.com/a-new-digital-world-is-emerging-thats-too-fast-for-us-1286428447.
[29] M. Eisen, "Amazon's $23,698,655.93 book about flies", It is NOT Junk, 22 April 2011. Available from http://www.michaeleisen.org/blog/?p=358.
[30] TechDirt, "The Infinite Loop Of Algorithmic Pricing On Amazon ... Or How A Book On Flies Cost $23,698,655.93", 25 April 2011. Available from https://www.techdirt.com/articles/20110425/03522114026/infinite-loop-algorithmic-pricing-amazon-how-book-flies-cost-2369865593.shtml.
[31] System of systems is can be understood as a "collection of task-oriented or dedicated systems that pool their resources and capabilities together to create a new, more complex system which offers more functionality and performance than simply the sum of the constituent systems." For more information, see https://en.wikipedia.org/wiki/System_of_systems. See also UNIDIR, *The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches. A primer*, UNIDIR Resources no. 6, 2017. Available from bit.ly/UNIDIR_AWSPrimer.

well as whether there may be unique or particularly acute vulnerabilities in autonomous weapon systems. These two categories are briefly described below.

Exacerbating vulnerabilities already seen in conventional weapon systems

With Stuxnet, the world witnessed how offensive cyber operations can shut down, damage or take control over physical systems or objects.[32] Following a review of US weapon systems in 2014, Michael Gilmore, the then-Director of the above-mentioned DOT&E Directorate, concluded that nearly all the tested systems were significantly vulnerable to cyberattack, and new vulnerabilities were still being discovered.[33]

Gilmore's annual reports identify specific vulnerabilities in software as well as in communication links. The seriousness of cyber vulnerabilities in existing systems and those already in development is so significant that it likely influenced President Trump's choice for nomination for Gilmore's successor, cyber expert Robert Behler. The team behind Mayhem, the winner of the 2016 Grand Cyber Challenge (see Box 1) was recently awarded a contract by the United States Department of Defense in "an effort to find coding flaws in both operating systems and custom programs used by the U.S. military".[34]

The Chairman's Food-for-Thought paper, prepared for the 2017 GGE on LAWS, asked whether "autonomous machines [can] be made foolproof against **hacking**".[35] The answer from industry experts appears to be a clear no.[36] While weapon systems can and must be hardened, it is unlikely that there can ever be a system that is completely invulnerable. There is no reason to believe that increasingly autonomous weapon systems will be immune to the vulnerabilities identified in traditional weapon systems.

A software vulnerability may be replicated throughout all weapons of the same class. While this vulnerability exists in traditional weapon systems, **systemic vulnerabilities** in an increasingly autonomous weapon system might go unnoticed for longer periods, as human operators are further removed from the system in space and in time, and their ability to interact with it is reduced.

Countermeasures to ICT-dependent weapon systems might include techniques such as **jamming** communications.[37] An **electromagnetic pulse (EMP)** weapon could also cause widespread technology failures, possibly even in hardened systems. Electronic warfare is beyond the scope of this paper, but it points to a potential systemic vulnerability—a brittleness—in these technologies.[38]

---

[32] While Stuxnet attack on Iran's Natanz Nuclear Facility, which destroyed numerous nuclear centrifuges by causing them to spin out of control, is perhaps the most well know, it is not the only instance of a cyber operation with kinetic effect.

[33] See Andrea Shalal, "Nearly every U.S. arms program found vulnerable to cyber attacks", *Reuters*, 21 January 2015. Available from http://www.reuters.com/article/us-cybersecurity-pentagon/nearly-every-u-s-arms-program-found-vulnerable-to-cyber-attacks-idUSKBN0KU02920150121. In his annual reports, the DoT&E Director has focused increasingly on how to defend cyberattacks and vulnerabilities in weapons platforms. See, for example, his most recent report, available from http://www.dote.osd.mil/pub/reports/FY2016/pdf/other/2016DOTEAnnualReport.pdf.

[34] The mission of Project Voltron is "leveraging breakthrough artificial intelligence in order to discover issues in military software". See https://www.cyberscoop.com/mayhem-darpa-cyber-grand-challenge-dod-voltron/.

[35] See Group of Governmental Experts of the High Contracting Parties to the 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, *Food-for-Thought Paper Submitted by the Chairperson (advance version),* 4 September 2017, p. 2.

[36] For instance, the principal data scientist at the cyber security company Endgame recently stated: "Machine learning security is not foolproof". See Alfred Ng, "Stop cyberattacks. Just add robots", *CNET*, 1 September 2017. Available from https://www.cnet.com/news/cyberattacks-artificial-intelligence-ai-hackers-defcon-black-hat/.

[37] Some argue that this supports the need for increasingly autonomous weapon systems in order to have a more resilient arsenal.

[38] There is growing interest in EMP weapons as an effective means to neutralize an adversary's communications infrastructure. For example, see the recent tests of the Counter-electronics High-powered Advanced Missile Project (CHAMP). Available from: http://mil-embedded.com/news/raytheon-emp-missile-tested-by-boeing-usaf-research-lab/.

Supply chain vulnerabilities are another source of concern. Much of the innovation in increasingly autonomous technologies is from the private sector, with the military adopting "spin on" applications. There are risks associated with integrating civilian technologies into military systems where military-level security has not been designed into all of the components and is rather "retrofitted" onto them. All privately sourced or commercially developed hardware and software components represent potential vulnerabilities. These vulnerabilities can be included deliberately (such as by inserting backdoors in the code) or unintentionally (such as errors or weaknesses in the code itself).

Potential cyber vulnerabilities unique to—or particularly acute in—AWS

Due to their nature, physical weapon systems with increasingly autonomous features will have some vulnerabilities to cyber operations that are unique or particularly acute.

Were increasingly autonomous weapons to operate **in a communications denied environment, how would operators know that the weapon system was working correctly and had not been compromised**? Zero-day exploits are of particular concern. Consider that the "Heartbleed" bug "existed in many of the world's computer systems for nearly *two and a half years*, for example, before it was discovered … . Analysts have estimated that, on average, such flaws *go unremediated for 10 months* before being discovered and patched, giving nefarious actors ample opportunity to wreak havoc in affected systems before they move on to exploit new terrain."[39] [emphasis added]

Increasingly autonomous weapons with long loiter or deployment times also raise concerns. The longer an autonomous object is deployed, particularly if it is out of communication, the more time an adversary has to discover and exploit vulnerabilities undetected. If a security vulnerability were identified in an autonomous long-loiter object operating in environments with limited communications, would there be an ability to patch the vulnerability remotely, recall the object, or at a minimum initiate a fail-safe shutdown mode?

Opaque systems and explainability. In many ways, AI delivers results that are currently incomprehensible to its designers. We simply do not have the tools to understand *how* it came to its answer. Movements toward "explainable AI" are underway but are far from providing serious comprehension to human operators.

As systems become increasingly autonomous and, concurrently, humans become decreasingly present in their operation and oversight, there is a serious risk that humans may be unable to serve as a redundant safety feature. If a human is no longer capable of intervening, we have the designed the human out of the system—either in time or in ability to understand.

Returning to the example of stock market flash crashes, these crashes were due, in part, to the sheer speed of these high-frequency trading agent systems.[40] However they were also due to the opaqueness of the system and people's inability to foresee that possibility. This made it almost impossible in practical terms for humans to intervene in time to prevent these unforeseen consequences. A recent UNIDIR report highlighted that with more complex military systems, especially those with autonomous functions, it would be reasonable to expect new failure modes. Worse, a single failure could cause potentially broad, cascading effects that humans cannot detect

[39] DARPA, "Top Teams' Automated Cybersecurity Systems Preparing for Final Face-off", 2016. Available from https://www.darpa.mil/news-events/2016-07-13.

[40] Todd C. Frankel, "Mini Flash Crash? Trading Anomalies on Manic Monday Hit Trading Investors", *The Washington Post*, 26 August 2015. Available from https://www.washingtonpost.com/business/economy/mini-flash-crash-trading-anomalies-on-manic-monday-hit-small-investors/2015/08/26/6bdc57b0-4c22-11e5-bfb9-9736d04fc8e4_story.html.

or stop in real-time.[41] These risk factors are equally valid for intentional manipulation or targeting of the weapon system via cyber means.

The benefits and risks of "unexpected moves". Even if AWS were observed or monitored by their operators (such as an "on the loop" system where a human could intervene if necessary), machine learning systems sometimes behave in unexpected or unanticipated ways—this is part of the power of AI. These unexpected moves are not necessarily bad in themselves—they can simply represent creative problem solving. When AlphaGo made a particularly surprising choice in its 2016 match against Lee Sedol, experts called it a "non-human" move, a strategy previously unimaginable until AlphaGo did it.[42] Harnessing AI to solve problems in previously unimagined ways is ultimately the desired benefit of AI. At the same time, it means that as operators become accustomed to AI-enabled objects behaving in surprising ways, they would have one less metric to use to determine if a weapon system with autonomous features was behaving in a way other than intended (whether due to internal error or malicious intervention)—thereby further diminishing the observer's ability to intervene.

Machine learning in adversarial settings. Perhaps the most important interaction between autonomous weapon systems and cyber operations would be to use machine learning in order to produce an undesirable or even unlawful effect. There are numerous ways in which autonomous weapon systems could be subverted via cyber operations. Sensors can be spoofed or tricked in ways that are not visible to humans. Data integrity can be sabotaged, training data sets tampered with, and data streams corrupted.

> *In an adversarial environment, such as in war, enemies will likely attempt to exploit vulnerabilities of the system, whether through hacking, spoofing (sending false data), or behavioral hacking (taking advantage of predictable behaviors to "trick" the system into performing a certain way). While any computer system is, in principle, susceptible to hacking, greater complexity can make it harder to identify and correct any vulnerabilities.*[43]

A well-known scenario used to describe this problem is that of autonomous vehicles recognizing traffic signs. Exposed to an image of a sign, a neural network classifies it in one of its predefined classes of signs, identifies it as a stop sign and the vehicle acts on that conclusion. If an adversary were to alter the input in such a way that the neural net misclassifies the stop sign as another sign—such as a yield or speed limit sign—the vehicle would be misled into taking the wrong action. Increasingly autonomous systems will be operating in adversarial environments where opponents will be seeking to subvert machine learning in malicious ways, such as misclassification of target sets. In such an environment, adversarial samples—"carefully modified inputs crafted to dictate a selected output"—would be designed to "force a target model to classify them in a class different from their legitimate class—for instance spam emails that bypass the spam filter. … In general adversaries want to perturb the sample as little as possible so that to a human observer, for example, it remains indistinguishable from the original unaltered sample."[44]

---

[41] UNIDIR, *Safety, Unintentional Risk and Accidents in the Weaponization of Increasingly Autonomous Technologies*, UNIDIR Resources no. 5, 2016. Available from http://www.unidir.org/files/publications/pdfs/safety-unintentional-risk-and-accidents-en-668.pdf.

[42] For a deeper exploration of the non-human move (Move 37), see Cade Metz, "How Google's AI Viewed the Move No Human Could Understand", *Wired*, 14 March 2016. Available from https://www.wired.com/2016/03/googles-ai-viewed-move-no-human-understand/.

[43] Paul Scharre, "Autonomous Weapons and Operational Risk", Center for a New American Security, 2016. Available from https://s3.amazonaws.com/files.cnas.org/documents/CNAS_Autonomous-weapons-operational-risk.pdf?mtime=20160906080515.

[44] Patrick McDaniel et al., "Machine Learning in Adversarial Settings", *IEEE Security and Privacy*, May-June 2016, pp. 68–72. Available from https://pdfs.semanticscholar.org/a69e/5b952a3d334d88555ad767b29ddd52d67cad.pdf.

Imagine a scenario where an autonomous weapon system was subverted to attack critical infrastructure, civilians or even turned against those who deployed it or their allies (intentionally causing a "friendly fire" incident). Were an adversary able to successfully carry out such an attack, the effects would be immediate and long lasting: in addition to the physical destruction, there would be a debilitating loss of trust in autonomous systems. A whole class of weapon systems might be "grounded" as a result while an investigation was conducted to see if the event was a malfunction, human error or an adversarial attack.

This report has mainly focused on cyber vulnerabilities in "autonomy in motion" systems. However, "autonomy at rest" systems, such as decision support aids used by the military, could also be vulnerable to cyber exploits. Militaries, for example, are integrating machine learning to process massive amounts of incoming sensor data, both in order to relieve human analysts and to gain efficiencies. Such an autonomy at rest system could be vulnerable to manipulation in the manners described above.[45]

[45] Project Maven, part of the recently established "Algorithmic Warfare Cross-Functional Team" overseen by United States Undersecretary of Defense for Intelligence, is one example: "Project Maven focuses on computer vision—an aspect of machine learning and deep learning—that autonomously extracts objects of interest from moving or still imagery." For more information, see https://www.defense.gov/News/Article/Article/1254719/project-maven-to-deploy-computer-algorithms-to-war-zone-by-years-end.

IV. Conclusions

How can the nature of increasingly autonomous intelligent agents be best reflected in the CCW discussions? As the GGE on LAWS starts its deliberations, High Contracting Parties to the CCW will need to **consider whether and how increasingly autonomous cyber technologies fit into the definitional discussion**. Some may think that cyberattacks are not "lethal" and therefore fall outside the framing of the issue within CCW. However, as seen with Stuxnet, cyber operations can have kinetic effects, and it is possible that such effects could have direct lethal consequences.

Depending on how AWS are defined, cyber operations might be included in the ongoing discussions. The Chairman's Food-for-thought paper to the 2017 CCW GGE asks whether autonomous systems are best visualized as physical robots or virtual machines.[46] On the one hand, the International Committee of the Red Cross (ICRC) has proposed a working definition of an autonomous weapon system: "Any weapon system with autonomy in its critical functions. That is, a weapon system that can select and attack targets without human intervention."[47] This definition would not, *a priori*, exclude autonomous cyber operations. On the other hand, the Autonomy Directive of the United States Department of Defence explicitly states that the Directive does not apply to "autonomous or semi-autonomous cyberspace systems for cyberspace operations".[48]

Even if an explicit decision is taken to exclude cyber from the CCW discussion, as the above-mentioned Directive has done, policy-makers will need to remain aware of the interactions between these two domains. In addition, High Contracting Parties should be mindful that their discussions on human control and judgement in relation to autonomous systems could have normative or even legal ramifications for increasingly autonomous cyber operations.

Awareness of potential cyber vulnerabilities must be paramount in the consideration of the risks of increasingly autonomous weapon systems, including an awareness about the range of States and non-state actors with the ability and motivation to exploit these vulnerabilities. Machine learning and artificial intelligence relies on software. One cannot discuss the safe exploitation of AI without also addressing cyber security—any cyber vulnerability is a gateway to a much more powerful level of processing and potential decision making. **Cyber fault management simply has higher stakes with autonomous systems**.

In order to effectively address the weaponization of increasingly autonomous technologies, **governments are likely to need to draw upon more computer science and AI expertise** to ensure that the intangible "soft" components of increasingly autonomous systems are understood and addressed in an appropriate and realistic way.

This brief paper points to some of the limitations of discussing increasing autonomy in weapon systems if the physical object of the weapon is overemphasized, and its intangible components, such as its software, are underappreciated. Considering the similarities and interplay between increasing autonomy in both virtual and physical systems can produce insights of value to each distinct discussion. That said, **the international discussion on LAWS and that on the international security dimensions of cyber should <u>not</u> be merged.** There are many elements of the cyber discussion that are not applicable to the LAWS discussion, including how intangible operations challenge

[46] See *Food-for-Thought Paper Submitted by the Chairperson of the CCW GGE (2017), op. cit*.

[47] International Committee of the Red Cros (ICRC), *International humanitarian law and the challenges of contemporary armed conflicts*, Report to the 32nd International Conference of the Red Cross and Red Crescent held 8-10 December 2015 (published October 2015), pp 44-47. Available from https://www.icrc.org/en/download/file/15061/32ic-report-on-ihl-and-challenges-of-armed-conflicts.pdf.

[48] United States Department of Defense, "Autonomy in Weapons Systems", Department of Defense Directive 3000.09, section 2.b, 21 November 2012. Available from http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf.

fundamental concepts of the UN Charter such as “territory” and “armed attack”. Physical weapon systems do not pose these difficulties.

It isn’t hyperbole to say that machine learning breakthroughs are occurring at a pace that was unimaginable even a year ago. Some of these have immediate and obvious military applications.[49] Learning systems that can operate in adversarial environments and unfamiliar situations, using skills such as prediction and deception, are certainly militarily desirable—and some will point to these successes as justification for moving forward with increasingly autonomous weapons. However, evidence of progress in competitions and games—which are, however sophisticated they seem to us, limited environments in which the programmers have set many boundaries—should not give us false confidence that similar results or behaviour will occur in unconstrained environments where adversaries will be attempting to exploit vulnerabilities in weapon systems, as well as developing their own strategies that may fall outside the models or training of learning systems.

[49] In addition to the example of the Grand Cyber Challenge described in Box 1, consider that for example, in August 2017, Open AI fielded a bot player in the online game of Dota. “Success in Dota requires player to develop intuitions about their opponents and plan accordingly. [The Dota bot] has learned—entirely via self-play—to predict where other players will move, to improvise in response to unfamiliar situations, and how to influence other player’s allied units to help it succeed”. See https://blog.openai.com/dota-2/. For more detail, see also https://arstechnica.com/gaming/2017/08/ai-bot-takes-on-the-pros-at-dota-2-and-wins/.

UNIDIR

The Weaponization of Increasingly Autonomous Technologies:

Autonomous Weapon Systems and Cyber Operations

International discussions about autonomous weapons have thus far focused predominantly on conventional weapon systems. These systems are not, however, the only domain in which technological developments in autonomy can have an impact on international security. Rapid advances in machine learning and artificial intelligence also have a significant impact in the field of cyber security, and in particular for offensive operations carried out in cyberspace, so-called "cyber operations". As this paper explains, the interaction of cyber operations and increasingly autonomous physical weapon systems may give rise to new security challenges, as these interactions can multiply complexity and introduce new vulnerabilities.

No. 7

UNIDIR RESOURCES