Ethics and Information Technology (2023) 25:7
https://doi.org/10.1007/s10676-023-09679-w

ORIGINAL PAPER

Legal and ethical implications of autonomous cyber capabilities: a call for retaining human control in cyberspace

Marta Stroppa[1]

Published online: 9 February 2023

Keywords Autonomous cyber capabilities · Artificial intelligence · Cyberspace · Human control

Introduction

Artificial Intelligence (AI) has multiple military applications in all the operational domains, including cyberspace. Among other things, AI is increasingly used to augment the level of autonomy of both defensive and offensive cyber operations, to the extent that AI-enabled cyber capabilities will be able to work without real-time human intervention.

On the one hand, the development and deployment of autonomous cyber capabilities might be seen as operationally desirable, as they may outperform human soldiers, operating at a speed and scale that is beyond human control. For this reason, multiple States have started researching and developing these technologies. At the same time, however, autonomous cyber capabilities risk to be also highly unpredictable, unreliable and unexplainable, especially when they are deployed in hostile and dynamic scenarios. As such, their potential use in the military domain raises significant legal and ethical concerns.

To this date, however, discussions about cyber capabilities, AI and autonomous systems have proceeded on different tracks.[1] Although the Open-ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security and the Group of Governmental Experts on Lethal Autonomous Weapons Systems have respectively acknowledged their potential overlaps, none of them have explored the threats that might arise from the use of AI-enabled autonomous weapons in the virtual domain.[2] Similarly, in the Tallinn Manual 2.0, the International Group of Experts has acknowledged the capacity for autonomous operations of software agents and worms when defining those terms, but it did not explore their legal and ethical implications.[3] This is counterintuitive, especially if we consider that the cyberspace might well be the first area where autonomous weapons will be consistently deployed, given that it is easier and cheaper to introduce AI-enabled autonomy in the virtual domain.

This paper aims at offering a preliminary analysis of the challenges that might arise while using AI-enabled autonomous cyber capabilities in the use of force and conduct of hostilities, with the hope of contributing to future debates on the matter. After outlining the increasing use of AI in cybersecurity and its advantages, the paper will analyze the main legal and ethical challenges deriving from the use of autonomous cyber capabilities in war. It will be argued that autonomous cyber capabilities are unlikely to operate according to international law and ethical values, especially when they are used in complex scenarios. As such, it will be suggested that, in order to promote a responsible use of AI in cyberspace for military purposes, a possible way out of the above-mentioned concerns might be that of exercising a context-based degree of human control on autonomous cyber capabilities.

[1] Rain Liivoja and Ann Väljataga, *Autonomous Cyber Capabilities under International Law* (Tallinn: NATO CCDCOE, 2021), page 1. See also UNIDIR, "The Weaponization of Increasingly Autonomous Technologies: Autonomous Weapon Systems and Cyber Operations," UNIDIR Resources, 2017, pages 1–3.

[2] Liivoja and Väljataga, *Autonomous Cyber Capabilities under International Law*, pages 1–2.

[3] Michael N. Schmitt, ed., *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, 2nd ed. (Cambridge University Press, 2017), pages 567–568.

✉ Marta Stroppa
marta.stroppa@santannapisa.it

[1] Ph.D. Candidate at the Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà 33, 56127 Pisa, PI, Italy

Towards an increasing use of AI in cybersecurity

Rapid advances in AI are having a significant impact in the field of cybersecurity, especially in the military domain.[4] The potential use of AI in the digital sphere for both defensive and offensive purposes is indeed considerable, as it would allow States to further strengthen their networks' robustness, resilience and response against hostile cyber operations, providing strong tactical and strategic advantages.[5]

Significantly, the prospective of using AI techniques to increase the level of autonomy and adaptability of cyber capabilities is particularly attractive, as it would allow States to strengthen their offensive cyber capabilities while overcoming the deficiencies of conventional cybersecurity systems.[6] Furthermore, the use of AI-enabled systems in the digital domain may be even more advantageous than in the physical domain. First, there is no need to deploy human soldiers (or robots) on the territory of an enemy State, further reducing the risks to which human soldiers are exposed on the battlefield. Second, cyber capabilities do not require significant infrastructures, financing or physical space for development and deployment – hence, also the capital expenditure will be lower.[7] Third, autonomous cyber capabilities allow for a *dilatatio* of the temporal and geographical scope of war, since they are capable of self-reproducing and replicating all around the world for a potentially indefinite amount of time.[8]

Consequently, States are increasingly investing in the research and development of the so-called "autonomous cyber capabilities", namely those cyber capabilities that are able to "perform some task without requiring real-time interaction with a human operator".[9] Among them, France[10], Germany[11], the United Kingdom[12] and the United States[13] have all expressed their interest in developing such technologies. The Mayhem Cyber Reasoning System (CRS), perhaps the most notorious example of autonomous cyber capability, was precisely developed within the framework of the United States Defense Advanced Research Project Agency's 2016 Grand Cyber Challenge. Winner of the competition, the Mayhem CRS is a software capable of autonomously detecting external intrusions, identifying their origin, stopping the external intrusions, and exploiting the adversaries' software's vulnerabilities in order to harm the system where the intrusions originated from.[14]

While it has been argued the Mayhem CRS was significantly weaker than the performance of cyber security experts, this was one of the first attempts to fully automate cyber defense.[15] Despite it might be hard to reach fully autonomous cyber capabilities in the short term, several progresses have been made in the field.[16] To this day, there have been important developments in the automatization of threat and intrusion detection, as well as in the sharing of cyber security intelligence.[17] Such tools may be used by cyber security experts to support the decision-making process. Furthermore, it is important to consider that in the next years States are expected to further invest in the research and development of AI solutions for cybersecurity. Estimates indicate that the market for AI in cybersecurity for

[4] Amy Ertan et al., "Cyber Threats and NATO 2030: Horizon Scanning and Analysis" (Tallinn: NATO CCDCOE, 2020), page 90.

[5] Mariarosaria Taddeo, Tom McCutcheon, and Luciano Floridi, "Trusting Artificial Intelligence in Cybersecurity Is a Double-Edged Sword," *Nature Machine Intelligence* 1, no. 12 (2019): 557–60, page 557.

[6] Salvador Llopis Sanchez, "Artificial Intelligence (AI) Enabled Cyber Defence," *European Defence Matters*, no. 14 (2017): 18.

[7] Daniel Trusilo and Thomas Burri, "Ethical Artificial Intelligence: An Approach to Evaluating Disembodied Autonomous Systems," in *Autonomous Cyber Capabilities under International Law*, by Rain Liivoja and Ann Väljataga (Tallinn: NATO CCDCOE Publications, 2021): 51–66, page 58.

[8] *Ibid*, page 63.

[9] Rain Liivoja, Maarja Naagel, and Ann Väljataga, "Autonomous Cyber Capabilities under International Law" (Tallinn: NATO CCDCOE, 2019), page 10.

[10] On 18 January 2019, also the French minister of defence, Florence Parly, while introducing the French Military Cyber Strategy, insisted on "the future combination of cyber attacks and artificial intelligence, engaging in battle on networks at a speed that defies human understanding". See, in this regard, François Delerue, *Cyber Operations and International Law*, 1st ed. (Cambridge University Press, 2020), page 159.

[11] In 2021, Germany affirmed in its National Cyber Security its intention to continually examine "the opportunities for using AI systems to protect (government) IT systems". See, in this regard, German Federal Ministry of the Interior, "Cyber Security Strategy for Germany 2021," 2021, page 47.

[12] In 2022, the United Kingdom included in its National Cyber Strategy its commitment to explore, together with the Alan Turing Institute, whether and to what extent AI can be used to detect certain types of cyber-attacks. See, in this respect, United Kingdom, "National Cyber Strategy 2022", 2022, page 101.

[13] In a March 2016 interview, the US Deputy Secretary of Defense Bob Work declared that "[w]e will not delegate lethal authority for a machine to make a decision. […] The only time we will … delegate a machine authority is in things that go faster than human reaction time, like cyber and electronic warfare". See, in this regard, Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (New York: W. W. Norton & Company, 2018), page 95.

[14] *Ibid*, pages 216–222.

[15] Tanel Tammet, "Autonomous Cyber Defence Capabilities," in *Autonomous Cyber Capabilities under International Law*, by Ann Väljataga and Rain Liivoja (Tallinn: NATO CCDCOE Publications, 2021): 36–50, page 49.

[16] For a comprehensive analysis of autonomous and automated cyber defence capabilities, see Tammet, "Autonomous Cyber Defence Capabilities".

[17] *Ibid*, pages 40–41.

both the public and private sector will grow from US§8.8 billion in 2020 to a US§38.2 billion by 2027.[18]

The increasing use of AI in cybersecurity, however, presents important downsides that need to be taken further into account. In particular, AI's lack of predictability, reliability and transparency raises important concerns as to whether AI-enabled autonomous cyber capabilities are *de facto* able to act accordingly to the deploying State's intent in any given circumstances of use.[19] Furthermore, AI lacks also of human judgement and contextual awareness: while AI-enabled autonomous systems might be capable of acting on the basis of quantitative data, they are unable to process qualitative elements or to understand the context in which they are operating and its possible future evolutions.[20] Whether AI is used to fully automize cyber capabilities or to support cybersecurity experts in the decision-making, its inherent characteristics raise important legal and ethical concerns that, if left unaddressed, could pose significant problems for the society.[21]

Legal implications of autonomous cyber capabilities

From a legal point of view, the use of autonomous cyber capabilities in the military domain raises important questions with respect to the international law regulating the use of force (*jus ad bellum*) and the international law of armed conflict (*jus in bello*), as well as to who shall be considered responsible for violations of international law by means of AI-enabled autonomous cyber capabilities.

As to *jus ad bellum*, since autonomous cyber capabilities can be used for both offensive and defensive purposes, it is crucial to understand whether they can be used in compliance with Articles 2.4 and 51 of the United Nations Charter, respectively concerning the prohibition of threat or use of force and the right to self-defence. According to the Tallinn Manual 2.0, in order to constitute a violation of Article 2.4, a cyber-attack shall have the same scale and effects of a kinetic attack amounting to use of force,[22] which shall be assessed on the basis of the level of harm inflicted and in the light of both quantitative and qualitative factors.[23] Since algorithms are unfitted to analyze qualitative elements, it is unlikely that autonomous cyber capabilities will be able to assess *per se* the intensity of the attack they are expected to launch. Thus, cyber capabilities should be programmed by human operators to reach the desirable intensity of use of force against a designated target.

In the same way, it is also unlikely that fully autonomous cyber capabilities will be able to assess whether a hostile cyber operation amounts to "armed attack" – defined as "the most grave forms of the use of force"[24] – in order to invoke the right to self-defense under Article 51. Moreover, the lack of human judgement and contextual awareness would make it also difficult for these technologies to comply with the principles of necessity and proportionality.[25] Thus, before deploying them, human operators should verify whether autonomous cyber capabilities are able to contain the effects of their response, so that their consequences do not exceed the force necessary to terminate the incoming armed attack. Finally, it should be further explored whether autonomous cyber capabilities are capable of attributing the hostile cyber operation to the launching actor: while they may be able to establish 'technical attribution' of an incoming cyber-attack, it might be possible that the real source of the attack is hidden by means of various techniques of obfuscation (e.g. spoofing).[26] Should this happen, the State operating

[18] Market Data Forecast, "Global AI in Cyber Security Market," January 2022, available at: https://www.marketdataforecast.com/market-reports/ai-in-cyber-security-market.

[19] This argument was originally made by the International Committee of the Red Cross (ICRC) with respect to autonomous weapons systems, but can easily be extended also to autonomous cyber capabilities. See, in this respect, ICRC, "Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control?," April 3, 2018, page 14.

[20] Daniele Amoroso and Guglielmo Tamburrini, "Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues," *Current Robotics Reports*, 2020.

[21] Mariarosaria Taddeo, "Three Ethical Challenges of Applications of Artificial Intelligence in Cybersecurity," *Minds and Machines* 29, no. 2 (June 2019): 187–191, page 188.

[22] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, Rule 69.

[23] *Ibid*. In order to help States determining the scale and effects of a cyber-attack, the Tallinn Manuals provide for eight factors, which rely on both quantitative and qualitative elements: (a) severity; (b) immediacy; (c) directness; (d) invasiveness; (e) measurability of effects; (f) military character; (g) state involvement; and (h) presumptive legality of the attack. Beyond these eight factors, depending on the circumstances, States may look also to other circumstances, such as the prevailing political context, the identity of the attacker and the nature of the target.

[24] Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America) (International Court of Justice, June 27, 1986), paragraph 191. The threshold of "armed attack" is even more ambiguous in the cyber domain. See, in this respect, Michael N. Schmitt, "The Use of Cyber Force and International Law," in *The Oxford Handbook of the Use of Force in International Law*, by Marc Weller, Oxford Handbooks in Law (Oxford: Oxford University Press, 2015): 1110–1130.

[25] Michael N. Schmitt, "Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention," in *Autonomous Cyber Capabilities under International Law*, by Rain Liivoja and Ann Väljataga (Tallinn: NATO CCDCOE Publications, 2021):126–151.

[26] In this respect, there is a longstanding debate as to whether States are entitled to respond in self-defence only when hostile cyber operations are legally attributed to States, as sustained by the International Court of Justice in the Nicaragua judgement, or also when they are attributed to non-State actors, as claimed in the Tallinn Manual 2.0.

the autonomous cyber capability risks using force against a third State that has been made to look like the originator of a malicious cyber operation – and this might lead to an unforeseeable escalation of conflicts.

As to *jus in bello*, it should be questioned whether autonomous cyber capabilities can act in compliance with the principles of distinction, proportionality and precautions. To this day, autonomous cyber capabilities may be able to identify easily discernible targets in uncluttered environment (e.g. Stuxnet was able to autonomously identify a very specific type of programmable logic controller in an air-gapped network), but they seem not to have yet the necessary situational awareness to comprehend the context in which they are operating and to foresee future evolutions, especially when they are used in complex scenarios (e.g. Internet). This analysis heavily relies on subjective elements that cannot be fully understood by autonomous systems.[27] This is particularly problematic in cyberspace, since most infrastructures are dual-use in nature – i.e. they are used at the same time by civilians and the military.[28]

Thus, it is unlikely that autonomous cyber capabilities will be able to comply with the principles of distinction and proportionality in complex scenarios without any form of real-time human intervention. Consequently, it should be questioned whether the duty to take active precautions in the launch of an attack would require commanders to not deploy autonomous cyber capabilities in those cluttered scenarios.[29] It might be argued, indeed, that human commanders should at least exercise some form of control over autonomous cyber capabilities when the principles of distinction and proportionality are at stake.[30]

Finally, questions arise also as to who should be considered responsible for a violation of international law deriving from the use of autonomous cyber capabilities. While it has been suggested that autonomous systems should be granted some form of legal personality[31], in international law there is neither State practice nor evidence of *opinio juris* to draw such conclusion.[32] On the contrary, States have agreed in relation to autonomous weapons systems that "[h]uman responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines".[33] The same might therefore be said of autonomous cyber capabilities.[34] Thus, responsibility shall be solely attributed to the actors involved in the development and use of these technologies, which include *inter alia* the deploying State, the commanders and soldiers operating the autonomous system or the business enterprise developing and programming it.[35] The lack of human control, however, jeopardizes such ascription, especially when these systems act unforeseeably, against the actor's will or intent.

Thus, a responsible use of AI in cyberspace would require States to exercise a certain degree of human control over autonomous cyber capabilities. From a *jus ad bellum* perspective, human control should be exercised in the definition of the intensity of use of force and in the assessment of when and how to respond to a hostile cyber operation in self-defence. From a *jus in bello* perspective, human control is crucial when autonomous cyber capabilities are used in complex scenarios where the principles of distinction and proportionality risk to be violated. Finally, from an accountability respective, the exercise of human control would

The use of autonomous cyber capabilities raises important concerns with respect to the first position, as they are unable to establish the legal attribution of a hostile attack to a State. Such evaluation relies on subjective elements that cannot be processes by algorithms. Thus, human control shall be exercised in the assessment of which is the responsible State of a hostile attack against which exercise the right to self-defence. *Ibid*, page 148.

[27] Daniele Amoroso, *Autonomous Weapons Systems and International Law: A Study on Human-Machine Interactions in Ethically and Legally Sensitive Domains* (Naples / Baden-Baden: Edizioni Scientifiche Italiane / Nomos Verlag, 2020), pages 60–70.

[28] In order to overcome this problem, it has been suggested to adopt a sort of "digital emblem" to identify objects and persons specifically protected from attacks (e.g. medical units, personnel and means), but it is neither clear how to implement it, nor whether autonomous cyber capabilities will be able to recognize them and exclude them from their attacks. See, in this respect, Felix E. Linker and David Basin, "Signaling Legal Protection during Cyber Warfare: An Authenticated Digital Emblem," *ICRC Blogpost*, September 21, 2021, https://blogs.icrc.org/law-and-policy/2021/09/21/legal-protection-cyber-warfare-digital-emblem/.

[29] Article 57 (2) (b) of Additional Protocol I to the Geneva Conventions.

[30] See, in this respect, Daniele Amoroso's interpretation of the principle of precaution with respect to autonomous weapons systems, in Amoroso, *Autonomous Weapons Systems and International Law*, pages 96–113.

[31] See, for instance, the European Union Parliament 2016 Report on Robotics, where it is suggested to "creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making food any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently". Mady Delvaux, "Motion for a European Parliament Resolution with Recommendations to the Commission on Civil Law Rules on Robotics" (Brussels: European Parliament, January 27, 2017), paragraph 59 (f).

[32] Rain Liivoja, Maarja Naagel and Ann Väljataga (Eds.), *Autonomous Cyber Capabilities under International Law*, page 32.

[33] Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System, Annex III to CCW/MSP/2019/9, Principle (b).

[34] Liivoja, Naagel, and Väljataga, "Autonomous Cyber Capabilities under International Law", pages 32–41.

[35] *Ibid.*

facilitate the ascription of responsibility whenever there is a violation of international law. The degree of human control should be context-specific: it should vary according to the context in which autonomous cyber capabilities are used, on the basis of both humanitarian and military considerations.[36]

Ethical implications of autonomous cyber capabilities

The use of autonomous cyber capabilities in the military domain also raises important questions from an ethical perspective. As underlined by the ICRC in its discussions over autonomous weapons systems, indeed, the concerns about the loss of human agency over weapons systems are not solely related to their compatibility with international law, but encompass also fundamental questions of acceptability to societal values.[37]

First, the fact that autonomous cyber capabilities might be more advantageous than other conventional weapons raises important concerns, as it might led to a proliferation of such technologies, with more and more States developing and deploying them for both offensive and defensive purposes. Questions about proliferation are also connected to autonomous cyber capabilities' self-replicating abilities, that allow them to multiply and propagate without direct human control.[38] Autonomous cyber capabilities' acceptability is strongly related to the context in which they are used. Therefore, the more autonomous cyber capabilities are able to self-replicating and propagating without human control, the more they are likely to be considered unacceptable.[39] The potentially limitless temporal and geographical scope of intervention of autonomous cyber capabilities creates indeed high uncertainty as to when and or where an attack will occur.[40] This may lead to an escalation of conflicts, resulting in flash cyber-war that would jeopardize the whole humankind.[41]

Secondly, the fact that it is possible to entirely delegate to computers the decision on whether to attack (or respond in self-defense against) another State raises important questions as to whether computers are more (or equally) fit than humans in taking such decisions, as they lack of the situational awareness and political understanding of reality.[42] After all, the use of force against another State and the exercise of the right to self-defense are political decisions taken in the framework of international law. Thus, it is unlikely that autonomous cyber capabilities will be able to take such decisions on their own. For this reason, delegating to computers the decision of using force is highly problematic.[43]

Thirdly, there are also concerns linked to the delegation of life-and-death decisions to computers, which would result in a loss of human dignity.[44] As argued with respect to autonomous weapons systems, indeed, "to allow machines to determine when and where to use force against humans is to reduce those humans to objects; they are treated as mere targets".[45] Since "[m]achines lack of morality and mortality"[46], delegating to a computer the decision to kill a human being undermines the very human dignity of those targeted, regardless of whether they are lawful targets under *jus in bello* or not.[47] This is even more problematic if we consider the unpredictability, unreliability and unexplainability of autonomous cyber capabilities.[48]

Finally, the removal of the human agency from autonomous cyber capabilities further weakens the moral responsibility of who launched the attack. Since autonomous cyber capabilities are still highly unreliable and unpredictable, the commanders' intent is not always reflected in the attack's outcome and consequences. As such, commanders and operators are unlikely to feel morally responsible for violations of international law resulting from autonomous cyber capabilities, especially when these are unforeseeable. On the contrary, they will often feel legitimized to shift their moral responsibility to the computer, which is perceived as a legitimate authority (or "moral buffer").[49]

[36] See, in this respect, the normative model of Meaningful Human Control elaborated by Daniele Amoroso with respect to autonomous weapons systems in Amoroso, *Autonomous Weapons Systems and International Law*, pages 217–260.

[37] ICRC, "Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control?", page 1.

[38] Trusilo and Burri, "Ethical Artificial Intelligence: An Approach to Evaluating Disembodied Autonomous Systems", page 62.

[39] ICRC, "Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control?", page 17.

[40] Trusilo and Burri, "Ethical Artificial Intelligence: An Approach to Evaluating Disembodied Autonomous Systems", page 63.

[41] Mariarosaria Taddeo, "Three Ethical Challenges of Applications of Artificial Intelligence in Cybersecurity," *Minds and Machines* 29, no. 2 (June 2019): 187–191, page 188; Guglielmo Tamburrini, "The AI Carbon Footprint and Responsibilities of AI Scientists," *Philosophies* 7, no. 1 (January 5, 2022): 1–11, page 8.

[42] Trusilo and Burri, "Ethical Artificial Intelligence: An Approach to Evaluating Disembodied Autonomous Systems", page 64.

[43] ICRC, "Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control?", pages 7–8.

[44] *Ibid*, pages 10–11.

[45] Christof Heyns, "Autonomous Weapon Systems: Human Rights and Ethical Issues" (Meeting of High Contracting Parties to the Convention on Certain Conventional Weapons, Geneva, April 14, 2016).

[46] Christof Heyns, "Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions to the Human Rights Council" (United Nations, April 9, 2013), paragraph 94.

[47] ICRC, "Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control?", pages 10–11.

[48] *Ibid*, pages 14–17.

[49] *Ibid*, pages 11–13.

Thus, also from an ethical perspective, a responsible use of AI in cyberspace would require States to exercise a certain degree of human control over autonomous cyber capabilities. Nonetheless, it is important to consider that even when human control is exercised, there is a risk human operators will place too much confidence on computers' results, trusting the machines more than their experience or judgement – this phenomenon is also known as "automation bias".[50] This is aggravated by the fact that human operators are often not fully aware of how autonomous cyber capabilities function, either due to a scarce training and to a lack of transparency in AI-enabled systems. Therefore, in order to be effective, human control should be informed - that is, human operators shall be trained to properly interface with the algorithm, which shall be comprehensible and explainable.[51]

Conclusion: a call for retaining human control in cyberspace

The several advantages that may derive from the use of AI in cybersecurity are leading many States to further research and develop autonomous cyber capabilities to be applied in the military domain for both defensive and offensive purposes. At the same time, however, the potential use of autonomous cyber capabilities raises important legal and ethical concerns that needs to be addressed.

In order to promote a responsible use of AI in cyberspace for military purposes, it will be suggested to retain a certain degree of human control over autonomous cyber capabilities, as it may help overcoming the main legal and ethical challenges that arise from the systems' unpredictability, unreliability and unexplainability, as well as from their lack of human judgement and situational awareness.

Of course, this paper acknowledges that cyberspace presents some hurdles that might jeopardize the exercise of human control over autonomous cyber capabilities, such as the very high speed at which cyber operations occur and the enormous quantity of transferred data. Nonetheless, it also contends that it is possible to overcome such challenges by exercising a context-specific human control that foresees different levels of human intervention, according to the context in which autonomous cyber capabilities are used, and that takes into account both humanitarian and military considerations.

As such, this paper hopes that future debates on autonomy, AI and cyberspace will take into consideration both the advantages and the risks related to their interplay, while further discussing possible solutions to promote a responsible use of AI for military purposes in cyberspace.

Funding Open access funding provided by Scuola Superiore Sant'Anna within the CRUI-CARE Agreement.

[50] Amoroso, *Autonomous Weapons Systems and International Law*, pages 238–239.

[51] *Ibid*, pages 246–249.