Cyber Weapons and Artificial Intelligence: Impact, Influence and the Challenges for Arms Control

Thomas Reinhold and Christian Reuter

Abstract As cyber weapons and artificial intelligence technologies share the same technological foundation of bits and bytes, there is a strong trend of connecting both, thus addressing the imminent challenge of cyber weapons of processing, filtering and aggregating huge amounts of digital data in real time into decisions and actions. This chapter will analyze this development and highlight the increasing tendency towards AI enabled autonomous decisions in defensive as well as offensive cyber weapons, the arising additional challenges for attributing cyberattacks and the problems for developing arms control measures for this "technology fusion". However, the article also ventures an outlook how AI methods can help to mitigate these challenges if applied for arms control measures itself.

1 Introduction

The idea of the weaponization of cyber tools has been under discussion for some time (Reinhold & Reuter, 2019b; Werkner & Schörnig, 2019). Many military or national security doctrines worldwide have adapted to the development that software can be designed, injected, triggered and controlled in foreign IT systems to perform tasks ranging from espionage to sabotage. This has been done from the perspective of necessary and appropriate defensive measures but also partly as a new category for offensive planning. Although no common international understanding has yet been reached on the threats posed by cyber weapons and their prevention, let alone a binding legal instrument, this field is already beginning to change due to the emergence of improved algorithms in artificial intelligence and machine learning (AI/ML) and their potential application for or against cyber weapons (Schörnig, 2018; US-DOD, 2018b). Given the fact that cyber and AI/ML measures are *natural siblings* from a technical perspective, the following text provides an assessment of

T. Reinhold (✉) · C. Reuter
Chair of Science and Technology for Peace and Security (PEASEC), Department of Computer Science, Technical University of Darmstadt, Darmstadt, Germany
e-mail: reinhold@peasec.tu-darmstadt.de; reuter@peasec.tu-darmstadt.de

T. Reinhold, N. Schörnig (eds.), *Armament, Arms Control and Artificial Intelligence*, Studies in Peace and Security, https://doi.org/10.1007/978-3-031-11043-6_11

how AI/ML methods could influence the development of malicious cyber activities based on an overview of their current state. Regarding the threats posed by this development for international security and new challenges for arms control, the text seeks on the one hand to assess how arms control approaches should prepare for AI/ML-driven cyber weapons. On the other hand, the text also examines the question of whether and how this technology can improve arms control approaches combating the weaponization of cyberspace.

2 Cyber Weapons and the Militarization of Cyberspace

Technological and scientific advances, especially the rapid evolution of information technology (IT), play a crucial role in questions of peace and security (Reuter, 2019). First and foremost, the most significant impact of the discussions and developments regarding the weaponization of cyberspace in recent years has been on its influence and the changes it has introduced to national and international security doctrines. An important incident has been the discovery of Stuxnet (Langner, 2013), malware developed by the US and Israel (Nakashima & Warrick, 2012) and targeted against a specific nuclear enrichment facility in Iran. Stuxnet manipulated the industrial control system of the facility by covertly changing thresholds and parameters of the control software to sabotage the enrichment process. This highly specified and *hand-crafted* attack on IT systems forced state leaders and decision-makers to recognize the vulnerabilities in computer systems and the threat that arises from the high degree of dependency on IT in economic, societal and government sectors. Especially critical infrastructures are now perceived to be high-risk targets for state and non-state cyberattacks. Although this was not the first cyber incident, and was hardly news for IT security specialists, the Stuxnet event demonstrated the technological possibility of crossing the cyber-physical barrier with dedicated malware and showed how to carry out actual physical destruction (Symantec, 2013) by remotely accessing and altering software. It also revealed the intent and the capacities of certain nation-states to develop and deploy such measures. In recent years states have reacted to this development by developing defensive measures to protect national IT infrastructures, extending national security and military doctrines to provide legal and organizational frameworks and establishing new and dedicated government or military institutions for these tasks. In addition, a large number of countries have also adopted offensive strategies, included those involving cyberspace, in their military planning and have established human and technological capacities (UNIDIR, 2013). This situation was emphasized by similar announcements by different states such as the US (US-DOD, 2018a) and the United Kingdom (UK Government, 2016). In 2016, NATO also declared (NATO, 2016) that incidents involving matters of or in cyberspace could invoke application of Article 5 of the Washington Treaty and prompted its member states to establish necessary military cyber capacities able to defend the alliance in this domain. A further major development was the US adoption of a new *defend forward* cyber security strategy in

2018 (US-DOD, 2018a). Declaring the ineffectiveness of defending the national IT systems by establishing IT security measures for them, the new strategy shifts activities outward to focus on the IT systems of potential adversaries and establishes a persistent engagement of cyber forces. Constant activities within foreign IT systems should, according to the strategy, provide early warning of looming attacks and keep foreign cyber forces busy enough to prevent and deter cyberattacks in the first place (Healey, 2019).

2.1 The Current Situation of State-Driven Cyberattacks

When it comes to the application of cyber measures in actual physical warfare, however, it seems that cyberattacks more often play a supporting role in military conflicts and are currently not used for massive destruction but rather for reconnaissance as well as the gathering of combat-relevant information. Most of the known cyber incidents were either cases of espionage, campaigns for political influence (Desouza et al., 2020), targeted minor IT systems or were performed with valid user credentials for critical IT systems gathered via social engineering and classic intelligence work. Although the potential for massive destruction was suspected in some cases, only a few cases with explicitly designed and deployed destructive cyber weapons have been identified so far, such as *Shamoon* (SecureList, 2012) or *TRITON* (Miller et al., 2019), both of which were deployed to sabotage central IT systems of Saudi Arabian petrochemical companies. From a strategic perspective, malicious cyber tools seem to have become widely accepted as an additional measure in hybrid conflicts or similar situations that deliberately stay below the threshold of full-fledged military confrontation. The relatively inexpensive creation of offensive cyber capacities—compared with traditional armament—also empowers new international actors. For instance, the Democratic People's Republic of Korea (North Korea) has become a relevant actor in cyberspace and has been responsible for different incidents over the last years (Ji-Young et al., 2019) such as the hacking attacks against a subsidiary of Sony, banks in Bangladesh or cryptocurrency marketplaces (US-DHS, 2020). Finally, the trend toward the stockpiling of vulnerabilities and exploits as the *base material* for cyber weapons raises new international threats. Undisclosed vulnerabilities in popular software not only provide possibilities for attacks by the withholding party but, conversely, leave anyone using the product vulnerable to attacks by any actor which becomes aware of the weak spot. The incidents of *WannaCry* (GReAT, 2017) and *NotPetya* (Mimoso, 2017), with their massive damage and commercial losses, are dramatic demonstration of this. Both malware campaigns exploited a vulnerability named *EternalBlue* that had been harbored and stockpiled by the US National Security Agency (Kubovic, 2018). The examples demonstrate on the one hand that states are increasingly developing and deploying offensive cyber capabilities, although trying to avoid serious damage to human life and staying below the threshold of IHL-prohibited aggressive actions. On the other hand, military cyber units are probably training and preparing for utilization of their capabilities in the event of conflicts. In addition, relatively cheap military cyber capabilities are revealing potential regional power shifts, thus increasing the probability of their application in smaller-scale conflicts.

3 How the Technology of Cyber Weapons and Its Application Will Evolve

A starting point for anticipating the influence and impact of AI/ML on the militarization of cyberspace, is the assessment of the possible evolvement of cyber weapons in general as well as consideration of future challenges regarding this type of technology. With the ever-growing automatization of all kinds of technological processes, IT systems are increasingly being integrated into physical systems and devices to control specific functions. Additionally, these IT systems will be further connected with each other (like the *Internet of Things*) and to cyberspace in order to perform tasks remotely (Russell, 2020). This means that defense against cyberattacks will involve an ever-increasing range of distributed digital devices that need to be made even more resistant against malicious influence, as well as chain effects due to interconnections and dependencies. In addition, with the increasing number of devices and the data they create, process or store, the amount of information that needs to be integrated and processed to detect anomalies and malicious operations will continue to rise. The range of possible attack vectors will further grow and diversify. Given the necessity to react to attacks in (almost) real time, the required decision-making must be accelerated and information processed almost instantly. This requires decision-making based on integrated mechanisms of autonomy or the filtering and pre-processing of information to compensate for the relative slowness and limited capacities of human operators (Burton & Soare, 2019). Moreover, this kind of automatization might possibly lead to a *cyber-vs-cyber* situation, where attacks are directly blocked by dedicated defensive measures without human intervention. Similar early consideration of offensive operations and an automatic infection of possible targets within cyberspace by an NSA-backed program called MONSTERMIND (Zetter, 2014) were exposed by Edward Snowden in 2013. Following the US *defend forward* and *persistent engagement* strategy, which will probably soon be adopted by other states, such developments will result in a further undermining of global IT security by means of the preparatory or precautionary installation of backdoors within foreign IT systems, in order to have the option of deploying the intended payload in time. As cyberspace is, on the one hand, the domain of military activities but, on the other hand, also represents the *physical space* that processes the transmission of any kind of action, the IT infrastructures, being its backbone, will obviously become relevant targets themselves. Finally, as the capability already exists, it is presumably only a matter of time until cyber capacities will be used and deployed openly in fully-fledged military conflicts, since situations already exist where the IT of military systems and weapons themselves have become targets (Perkovich & Hoffman, 2019).

4 How Artificial Intelligence and Machine Learning Could Influence Cyber Weapons

Reflecting on the possible impact of AI/ML on cyber weapons and the militarization of cyberspace, it is crucial to highlight that cyber and AI/ML measures are *natural siblings*. "*[AI and ML] share the idea of using computation as the language for intelligent behaviour*" (our italic) (Kersting, 2018). From a purely technological perspective, AI/ML is *just* software: algorithms based on complex computer code that can be integrated into decision processes. Hence, AI/ML is developed and deployed within the same domain as cyber tools and to a considerable extent requires similar know-how in programming, code logic and software life cycle management. In order to be effective, cyber tools must keep pace with the latest technological developments, software updates and the modernization of devices. To reach this level of adaptability and extendibility they are often based on modern development frameworks with modularized, extendable and interchangeable software architecture [see, for example, the *FLAME* malware platform (sKyWIper Analysis Team, 2012)]. Such architecture provides an ideal platform for an extension with AI/ML components. Additionally, computer code offers optimal conditions for creating and facilitating training and testing environments for military AI/ML applications, as the environment can be defined and shaped in every specific detail and according to the intended requirements. This reduces costs and the amount of research and development required. As described in the previous section, an important challenge for cyber as well as other military technologies is the growing amount of information that needs to be processed (Kersting & Meyer, 2018), in contrast to the decreasing time to react to incidents. This dilemma involves incidents within cyberspace but also situations where cyber tools facilitate the analysis of data and the processing of information in order to provide the basis for decision-making concerning physical systems such as weapons or reconnaissance systems. AI/ML algorithms, and especially modern approaches such as deep learning (Charniak, 2019), were developed specifically for cases involving processing large amounts of data, detecting patterns and filtering out relevant information from *digital noise*. According to Schörnig (2018), the "spectrum of possible applications [of AI in the military] ranges from the analysis of trade data to uncover clues for the proliferation of weapons of mass destruction, to the identification of landmines that is boosted by AI with improved ground penetrating radars." Because of such capabilities, military AI applications are likely to be integrated into cyber tools, as these usually have to deal with a large amount of digital data in trying to detect relevant patterns.

4.1 Explainability and Responsibility of AI-Enabled Cyber Weapons

An additional aspect of this development is that the automated conclusion process already mentioned and the resulting selection and decision about actions will be significantly changed when combined with AI/ML algorithms. Whereas the automatization of defensive cyber actions is hardly new, AI/ML are, in the sense of technology which produces an output for a given input without allowing reconstruction of the digital reasoning process or the *line of thought* of the machine or software that led to a specific decision. This creates situations in which the code produces decisions that are no longer deducible and thus prevent humans from intervening based on reasoning. When such AI/ML-enabled measures are used for offensive actions, this creates serious problems in connection with the necessary human integration and interaction (Schwarz, 2019). All these issues have already been the subject of heated debate in connection with autonomous weapon systems (AWS) regarding the responsibility and traceability of decisions (IPRAW, 2019; see the chapter from Anja Dahlmann). In order to address the problem of comprehensible AI/ML decisions, a dedicated field of research (XAI—Explainable Artificial Intelligence) (Gunning et al., 2019) is working on technical concepts that allow human operators either to follow the decisions during the reasoning process (ad-hoc XAI) or the decisions to be recapped once they are made (post-hoc XAI). So far, these approaches are mere theoretical concepts that lack general applicability and are hindered by specific technical features of machine learning such as the distributed and numerical representation of learned information (Barredo Arrieta et al., 2020). Additionally, it is questionable whether ad-hoc explainability can be used meaningfully in an environment characterized by extremely short response times, as the two conditions are mutually exclusive. The speed of reaction in combination with the black-box character of such tools may possibly prevent any opportunity for double-checking of decisions by human operators or for their intervention. Even if the code itself does not *pull the trigger*, human operators might tend to trust the decisions or pre-decisions of machines and follow their suggestions due to a lack of alternatives, time pressure or perceived lack of human influence or oversight (Bajema, 2019). As AI/ML algorithms are trained for specific situations and decisions before they are integrated into productive systems, the operators of the finished application might also be unlikely to know the specific details of the training data, nor have any chance to see, perceive or understand the assumptions and pre-conditions of this data. Besides, this inexplicability could lead to critical junctures in situations marked by high international tension. State actors on the brink of military conflict might lack the ability to communicate and explain automatically triggered actions or conclusions that led to their activities to other conflict parties, thus undermining a valuable measure of immediate conflict reduction. As unlikely as such a scenario currently seems, the discussion of application of AI/ML within the ongoing process of modernization of nuclear weapons arsenals (Field, 2019) is an example that highlights the consequences that are at stake (Boulanin, 2019). The application of AI/ML for militarized tools within cyberspace reveals an overall similarity to AWS (see the chapter from Anja Dahlmann). The debates on norms and limitations of the application of automated cyber tools could thus benefit from the lessons learned about the human role within the decision-making loop of technological systems and its consequences.

4.2 AI and the Pitfalls of the Attribution of Cyberattacks

The black-box character of AI/ML systems could also aggravate other features of cyberspace that are currently considered to be problematic, both in terms of the application of international humanitarian law (IHL) and of established norms of state conduct. One of these features of cyberspace concerns the attribution problem (Rid & Buchanan, 2015). Whereas the possibility of identifying attackers is essential for IHL and the states' right to use military force for self-defense (Grosswald, 2011), this task is complicated, time-consuming, and a forensic challenge due to the technical features of the cyberspace (Riebe et al., 2019). Digital information inherently contains a high degree of ambiguity and virtuality. Information can easily be copied, modified, or actively tailored to set false tracks. Consequently, the meaningfulness of information about cyber incidents needs to be critically evaluated to prevent false assumptions and reactions. Applying AI/ML measures to offensive operations will further reinforce this ambiguity and intensifies the problem of gaining a clear picture of what happened and identifying the actors behind it. The automatic AI/ML-driven evaluation of information about an incident inherently contains the problematic aspect of some conclusions about the origin of an attack being inadvertently misleading and the question of how to react proportionately. Such failure could be triggered either by incorrect or insufficiently trained algorithms, biased input information or by following intentionally created false trails (Herpig, 2019). Although the inner state of an AI is considered a *black box*, this condition is the result of the learning model and the data used to train the AI. Assuming that an attacker obtained knowledge of the model of an applied, static AI/ML and the data which had been used for its training—e.g., through leaks, reconnaissance, hacks, or insecure manufacturers' supply chains—it would be possible to replicate such an AI itself and thus calculate the output that this AI/ML would generate for a specific input. Such knowledge could enable an attacker to tailor its attacks either to avoid detection or to generate incorrect conclusions (Apruzzese et al., 2019). Finally, the development and application of AI/ML in commercial, non-military IT systems, especially in the field of IT security and automated network security surveillance and defense, will produce spill-over effects in military applications. This development will increase acceptance of such systems and put constant pressure on military decision-makers to deploy them to gain a supposed strategic or tactical advantage.

5 The Negative Impact on Arms Control of Artificial Intelligence in Cyber Weapons

The developments outlined above add to the existing challenges involved in applying stabilizing measures in security policy to cyberspace, such as working toward peace-sustaining cyber armament reduction and cyber arms control measures. Firstly, a general problem of cyberspace is its virtual character (Reinhold & Reuter, 2019a). Data has neither a specific geographic location nor a physical representation. It can be reproduced seamlessly and is not limited to a specific and unchanging location but can instead be distributed across different places, such as in cloud applications. As explained above in connection with the problem of data ambiguity, integrating an AI/ML system into existing cyber measures further increases aspects of virtuality and non-tangibility and thus undermines established concepts of arms control even more than software itself already does (Reinhold & Reuter, 2019c). Besides obvious dual-use problems (Riebe & Reuter, 2019), in practical terms the effortless duplication of digital data that concerns ready-made AI/ML applications as well as training data hinders the control of proliferation of military-grade AI/ML technology (see the chapter from Kolja Brockmann). This also negatively affects the ability to measure specific aspects of a regulated item, which is a core requirement of arms control (Burgers & Robinson, 2018). Like cyber tools in general, AI/ML algorithms are computer code, or even more abstractly, structured digital data. They are thus immune to any kind of countability and provide few starting points for measuring parameters that could provide meaningful classification or comparison with permissible thresholds. This missing feature also means a distinction between civil and military AI/ML systems that is capable of going beyond the mere declaration of the intended application cannot be made while also preventing any kind of classification of the capacity and performance of an AI/ML system. This situation constitutes a major obstacle to the development of viable verification approaches for AI/ML applications. Apart from that, as the performance of an AI/ML system depends to a large extent on its training, the question arises as to whether the trade and proliferation regulation of training data—either artificially, as tailor-made datasets or taken from real-life samples and situations—could provide a starting point for arms control and nonproliferation regimes. The chapter "Arms Control for Artificial Intelligence" in this volume on arms control for AI will further evaluate these possibilities and challenges.

6 How Can Artificial Intelligence Support Cyber Arms Control?

Apart from the challenges described above about how AI/ML algorithms can add to the already complicated cyber weapons debates and the attempts at peaceful development in this domain, such technologies could possibly also evolve into useful tools for cyber arms control and disarmament. In general, AI/ML algorithms are a good tool for combining and processing large amounts of different, heterogeneous, often noisy and rapidly changing data to detect patterns, regularities and *hidden* information (Lück, 2019). A specifically powerful aspect of this technology is the ability to identify similarities within data and find useful matching items that do not fully correspond to the trained items but relate to them with a high degree of certainty. This kind of detection quality is usually a problem that cannot be solved with hard-coded deterministic rules. By contrast, an AI/ML algorithm is able to identify relevant detection parameters during its training phase, establishing a self-developed filter for relevant and irrelevant information. As a result, AI/ML algorithms could prove to be the right tool for managing the information overload of IT systems (Kaufhold et al., 2020) and the challenge of finding the needle in the haystack. Such challenges could be the task of searching for anomalies in information provided by states in the context of confidence-building measures or processing surveillance imagery to detect military installations. A meaningful, currently unexplored application could be to control the proliferation of cyber weapons (Silomon, 2018) by monitoring the distribution and occurrence of specific parts of weaponized computer code. As already mentioned, code can easily be copied and will, in almost all cases, be slightly modified or extended to fit into existing cyber weapons, to work with the specific tools and programming frameworks, or to match specific target criteria. Any detection mechanisms searching for an exact piece of computer code will presumably fail to detect such modified versions. An AI/ML algorithm could be trained to circumvent this problem and to provide at least indicators and probability measures of whether and to what extent computer code matches a specific sample. A similar approach could be used to detect and identify actors behind cyberattacks. Even if this is not directly a task of arms control, it overlaps with the regulation of cyber weapons, because an actor is visible, detectable and identifiable by its behavior, by technical operations performed in foreign IT systems and by the tools employed (Sibi Chakkaravarthy et al., 2019). Whereas it is possible and common to counterfeit these indicators in order to lay a false trail, an AI could be used to detect unconscious similarities of the attackers' style, habits and methods. Institutionalized military cyber actors in particular develop their know-how and the required skills over time. They create, extend and modify their own toolsets and cyber weapon arsenals, which are then reconfigured, combined and adjusted for a specific operation (Olszewski, 2018). This means that specific actors often have *digital fingerprints* regarding their customary tools and hacking strategies. Nearly every cyber activity creates digital traces such as small pieces of code that attackers have previously used to perform their tasks, manipulate files, change system settings or log entries or IP addresses of remote IT systems where data has been copied. Such detectable traces are called *samples* and are already used to compare new code to known samples from prior incidents in order to draw conclusions about an alleged actor. Although captured samples like these rarely match existing samples perfectly, they do contain similarities as they come from the same complex cyber weapon project, use similar methods and approaches, or are more advanced versions of each other. Detecting these similarities and identifying cyber weapons is a task where AI/ML approaches and algorithms are highly suitable (Roberts, 2019). For example, such identification measures are already used by IT security forensics when analyzing cyber incidents (Kanzig et al., 2019). They are often combined with further indicators such as specific habits and ways of programming, the structuring of computer code or recurring phrases and names. Lastly, the black-box character of AI/ML applications could also be an advantage for arms control measures. An essential element of practical control and compliance monitoring of arms control regimes is the requirement that the actors involved do not want to disclose any sensitive information about the regulated or controlled item (Kütt et al., 2018). This requires technical procedures where participating parties—usually states—are required to disclose as little information as possible when verification is performed and verification devices are developed that conceal all processing steps. In addition, the participating parties would have to be convinced that the results will be reliable and trustworthy. Such a tool, in which a defined input leads to a binary decision of *is or is not a weapon*, could be achieved through AI/ML procedures. To prevent doubts regarding the reliability and the acceptability of the algorithm's decision it would be necessary to prevent any modification or tampering and to preserve the integrity of the algorithm and its trained state. This could be achieved by securing the AI/ML application with *digital seals*, cryptographically calculated unique values—usually very long numbers—like checksums and hashes that represent a specific state of arbitrary digital information. A recalculation of the digital seal would immediately reveal any modification as it would result in a different number if the information has been changed (Putz et al., 2019). These mere outlines of applicable approaches presumably have other peculiarities that need to be taken into account when it comes to real-world applications. Although this issue goes beyond the scope of this chapter, it shows that, despite new challenges, AI/ML approaches can also contribute to arms control.

7 Conclusion

This assessment has provided an overview of the possible development and impact of AI/ML methods on cyber weapons. It is based on current trends and technical AI/ML developments as well as on the already ongoing application of or research on AI/ML in other military fields of operation. The assessment shows that the military application of AI/ML for cyber related tasks will probably exacerbate an already tense situation involving a cyber arms race on the one hand and a lack of international measures to prevent destabilizing and harmful effects on the other. Established measures for arms control, whose application to cyber weapons is already hindered by specific technical features of these tools, will face further challenges. Furthermore, for military decision-makers AI/ML algorithms seem to provide solutions for enhancing their weapon systems and battlefield management capabilities through their ability to integrate, process and refine large amounts of digital data. This could provide a strong incentive for military decision-makers to pursue and apply these approaches. However, the assessment also showed that, in addition to the necessary questions of peace and conflict research regarding AI/ML in cyber weapons, technological developments reflect ongoing debates about lethal autonomous weapon systems. This makes it possible to participate in these discussions and to benefit from lessons learned. Finally, AI/ML approaches could also provide valuable insights into the challenges of arms control for cyber weapons and help to circumvent some of its technological pitfalls. Either way, artificial intelligence and machine learning are just beginning to find their way into military cyber systems, and the time has come to critically accompany this trend and conduct further research in order to promote peaceful development of cyberspace.

References

Apruzzese, G., Colajanni, M., Ferretti, L., & Marchetti, M. (2019). *Addressing adversarial attacks against security systems based on machine learning*. In 2019 11th International Conference on Cyber Conflict (CyCon) (pp. 1–18). https://doi.org/10.23919/CYCON.2019.8756865

Bajema, N. E. (2019, November 12). *Can humans resist the allure of machine speed for nuclear weapons?* Retrieved from https://outrider.org/nuclear-weapons/articles/can-humans-resist-allure-machine-speed-nuclear-weapons/

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Boulanin, V. (2019). *The impact of artificial intelligence on strategic stability and nuclear risk*. Retrieved from https://www.sipri.org/publications/2019/other-publications/impact-artificial-intelligence-strategic-stability-and-nuclear-risk-volume-i-euro-atlantic

Burgers, T., & Robinson, D. R. S. (2018). Keep dreaming: Cyber arms control is not a viable policy option. *Sicherheit und Frieden, 36*(3), 140–145. https://doi.org/10.5771/0175-274X-2018-3-140

Burton, J., & Soare, S. R. (2019). *Understanding the strategic implications of the weaponization of artificial intelligence.* In 2019 11th International Conference on Cyber Conflict (CyCon) (pp. 1–17). https://doi.org/10.23919/CYCON.2019.8756866

Charniak, E. (2019). *Introduction to deep learning*. The MIT Press. https://dl.acm.org/doi/book/10.5555/3351847

Desouza, K. C., Ahmad, A., Naseer, H., & Sharma, M. (2020). Weaponizing information systems for political disruption: The actor, lever, effects, and response taxonomy (ALERT). *Computers and Security, 88*, 101606. https://doi.org/10.1016/j.cose.2019.101606

Field, M. (2019, December 20). *As the US, China, and Russia build new nuclear weapons systems, how will AI be built in?* Retrieved from https://thebulletin.org/2019/12/as-the-us-china-and-russia-build-new-nuclear-weapons-systems-how-will-ai-be-built-in/

GReAT. (2017). WannaCry ransomware used in widespread attacks all over the world. *Securelist. Com*. Retrieved from https://securelist.com/wannacry-ransomware-used-in-widespread-attacks-all-over-the-world/78351/

Grosswald, L. (2011). Cyberattack attribution matters under article 51 of the U.N. Charter. *Brooklyn Journal of International Law, 36*(3), 1151–1181.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics, 4*(37). https://doi.org/10.1126/scirobotics.aay7120

Healey, J. (2019). The implications of persistent (and permanent) engagement in cyberspace. *Journal of Cybersecurity, 5*(1), 1–25. https://doi.org/10.1093/cybsec/tyz008

Herpig, S. (2019). *Securing artificial intelligence*. Retrieved from https://www.stiftung-nv.de/sites/default/files/securing_artificial_intelligence.pdf

IPRAW. (2019). *Focus on human control*. Retrieved from https://www.ipraw.org/wp-content/uploads/2019/08/2019-08-09_iPRAW_HumanControl.pdf

Ji-Young, K., Jong In, L., & Kyoung Gon, K. (2019). *The all-purpose sword: North Korea's cyber operations and strategies*. In 2019 11th International Conference on Cyber Conflict (CyCon) (pp. 1–20). https://doi.org/10.23919/CYCON.2019.8756954

Kanzig, N., Meier, R., Gambazzi, L., Lenders, V., & Vanbever, L. (2019). *Machine learning-based detection of C&C channels with a focus on the locked shields cyber defense exercise*. In 2019 11th International Conference on Cyber Conflict (CyCon) (pp. 1–19). https://doi.org/10.23919/CYCON.2019.8756814

Kaufhold, M.-A., Rupp, N., Reuter, C., & Habdank, M. (2020). Mitigating information overload in social media during conflicts and crises: Design and evaluation of a cross-platform alerting system. *Behaviour and Information Technology, 39*(3), 319–342. https://doi.org/10.1080/0144929X.2019.1620334

Kersting, K. (2018). Machine learning and artificial intelligence: Two fellow travelers on the quest for intelligent behavior in machines. *Frontiers in Big Data, 1*, 6. https://doi.org/10.3389/fdata.2018.00006

Kersting, K., & Meyer, U. (2018). From big data to big artificial intelligence? *KI – Künstliche Intelligenz, 32*(1), 3–8. https://doi.org/10.1007/s13218-017-0523-7

Kubovic, O. (2018). *One year later: EternalBlue exploit more popular now than during WannaCryptor outbreak*. ESET. Retrieved from https://www.welivesecurity.com/2018/05/10/one-year-later-eternalblue-exploit-wannacryptor/.

Kütt, M., Göttsche, M., & Glaser, A. (2018). Information barrier experimental: Toward a trusted and open-source computing platform for nuclear warhead verification. *Measurement, 114*, 185–190. https://doi.org/10.1016/j.measurement.2017.09.014

Langner, R. (2013). *To kill a centrifuge—A technical analysis of what Stuxnet's creators tried to achieve*. Retrieved from https://www.langner.com/wp-content/uploads/2017/03/to-kill-a-centrifuge.pdf.

Lück, N. (2019). *Machine learning powered artificial intelligence in arms control.* PRIF Report 8/2019. Retrieved from https://www.hsfk.de/fileadmin/HSFK/hsfk_publikationen/prif0819.pdf

Miller, S., Brubaker, N., Zafra, D. K., & Caban, D. (2019, April 10). *TRITON actor TTP profile, custom attack tools, detections, and ATT&CK mapping*. Retrieved from https://www.fireeye.com/blog/threat-research/2019/04/triton-actor-ttp-profile-custom-attack-tools-detections.html

Mimoso, M. (2017, June 28). New petya distribution vectors bubbling to surface. *Threatpost.Com*. Retrieved from https://threatpost.com/new-petya-distribution-vectors-bubbling-to-surface/126577/

Nakashima, E., & Warrick, J. (2012). Stuxnet was work of U.S. and Israeli experts, officials say. *The Washington Post*. Retrieved from https://www.washingtonpost.com/world/national-security/stuxnet-was-work-of-us-and-israeli-experts-officials-say/2012/06/01/gJQAlnEy6U_story.html

NATO. (2016). *Warsaw Summit Communiqué: Issued by the Heads of State and Government participating in the meeting of the North Atlantic Council in Warsaw 8-9 July 2016*. Retrieved from http://www.nato.int/cps/en/natohq/official_texts_133169.htm

Olszewski, B. (2018). Advanced persistent threats as a manifestation of states' military activity in cyber space. *Scientific Journal of the Military University of Land Forces, 189*(3), 57–71. https://doi.org/10.5604/01.3001.0012.6227

Perkovich, G., & Hoffman, W. (2019). From cyber swords to plowshares. In T. de Waal (Ed.), *Think peace: Essays for an age of disorder*. Last retrieved on 03.01.2022, from https://carnegieeurope.eu/2019/10/14/from-cyber-swords-to-plowshares-pub-80035

Putz, B., Menges, F., & Pernul, G. (2019). A secure and auditable logging infrastructure based on a permissioned blockchain. *Computers and Security, 87*, 101602. https://doi.org/10.1016/j.cose.2019.101602

Reinhold, T., & Reuter, C. (2019a). Arms control and its applicability to cyberspace. In C. Reuter (Ed.), *Information technology for peace and security—IT-applications and infrastructures in conflicts, crises, war, and peace* (pp. 207–231). Springer Fachmedien. https://doi.org/10.1007/978-3-658-25652-4_10

Reinhold, T., & Reuter, C. (2019b). From cyber war to cyber peace. In C. Reuter (Ed.), *Information technology for peace and security—IT-applications and infrastructures in conflicts, crises, war, and peace* (pp. 139–164). Springer Fachmedien. https://doi.org/10.1007/978-3-658-25652-4_7

Reinhold, T., & Reuter, C. (2019c). Verification in cyberspace. In C. Reuter (Ed.), *Information technology for peace and security—IT-applications and infrastructures in conflicts, crises, war, and peace* (pp. 257–275). Springer Fachmedien. https://doi.org/10.1007/978-3-658-25652-4_12

Reuter, C. (2019). Information technology for peace and security—IT-applications and infrastructures. In C. Reuter (Ed.), *Information technology for peace and security—IT-applications and infrastructures in conflicts, crises, war, and peace* (pp. 3–9). Springer Fachmedien. https://doi.org/10.1007//978-3-658-25652-4_1

Rid, T., & Buchanan, B. (2015). Attributing Cyber Attacks. *Journal of Strategic Studies, 38*(1–2), 4–37. https://doi.org/10.1080/01402390.2014.977382

Riebe, T., Kaufhold, M.-A., Kumar, T., Reinhold, T., & Reuter, C. (2019). Threat intelligence application for cyber attribution. In C. Reuter, J. Altmann, M. Göttsche, & M. Himmel (Eds.), *Science peace security '19—Proceedings of the interdisciplinary conference on technical peace and security research* (pp. 56–60). TUprints. Retrieved from https://tuprints.ulb.tu-darmstadt.de/9164/2/2019_SciencePeaceSecurity_Proceedings-TUprints.pdf

Riebe, T., & Reuter, C. (2019). Dual-use and dilemmas for cybersecurity, peace and technology assessment. In C. Reuter (Ed.), *Information technology for peace and security—IT-applications and infrastructures in conflicts, crises, war, and peace* (pp. 165–183). Springer Fachmedien. https://doi.org/10.1007/978-3-658-25652-4_8

Roberts, P. S. (2019, December 13). AI for peace. *War on the Rocks*. Retrieved from https://warontherocks.com/2019/12/ai-for-peace/

Russell, B. (2020). IoT cyber security. In F. Firouzi, K. Chakrabarty, & S. Nassif (Eds.), *Intelligent internet of things* (pp. 473–512). Springer. https://doi.org/10.1007/978-3-030-30367-9_10

Schörnig, N. (2018). Artificial intelligence in the military: More than killer robots. In B. Wolff (Ed.), *Whither artificial intelligence? Debating the policy challenges of the upcoming transformation* (pp. 39–44). Science Policy Paper des Mercator Science-Policy Fellowship-Programms.

Schwarz, E. (2019). Günther Anders in Silicon Valley: Artificial intelligence and moral atrophy. *Thesis Eleven, 153*(1), 94–112. https://doi.org/10.1177/0725513619863854

SecureList. (2012, September 11). *Shamoon the wiper: Further details (part II)*. Retrieved from https://securelist.com/shamoon-the-wiper-further-details-part-ii/57784/

Sibi Chakkaravarthy, S., Sangeetha, D., & Vaidehi, V. (2019). A survey on malware analysis and mitigation techniques. *Computer Science Review, 32*, 1–23. https://doi.org/10.1016/j.cosrev.2019.01.002

Silomon, J. (2018). Software as a weapon: Factors contributing to the development and proliferation. *Journal of Information Warfare, 17*(3), 106–123.

sKyWIper Analysis Team. (2012). *sKyWIper (a.k.a. Flame a.k.a. Flamer): A complex malware for targeted attacks*. Retrieved from https://www.crysys.hu/publications/files/skywiper.pdf

Symantec. (2013). *Stuxnet 0.5: The missing link*. Retrieved from https://docs.broadcom.com/doc/stuxnet-missing-link-13-en..

UK Government. (2016). *National cyber security strategy 2016–2021*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/567242/national_cyber_security_strategy_2016.pdf

UNIDIR. (2013). *The cyber index: International security trends and realities*. Retrieved from https://www.unidir.org/files/publications/pdfs/cyber-index-2013-en-463.pdf

US-DHS. (2020). *Guidance on the North Korean cyber threat*. Retrieved from https://www.us-cert.gov/ncas/alerts/aa20-106a.

US-DOD. (2018a). *National cyber strategy*. Last retrieved on 03.01.2022, from https://trumpwhitehouse.archives.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf

US-DOD. (2018b). *Summary of the 2018 Department of Defense AI strategy*. Last retrieved on 03.01.2022, from https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF

Werkner, I.-J., & Schörnig, N. (Eds.). (2019). *Cyberwar – die Digitalisierung der Kriegsführung*. Springer Fachmedien. https://doi.org/10.1007/978-3-658-27713-0

Zetter, K. (2014, August 13). *Meet monstermind, The NSA bot that could wage cyberwar autonomously*. Retrieved from https://www.wired.com/2014/08/nsa-monstermind-cyberwarfare/