Intelligent Environments 2018
I. Chatzigiannakis et al. (Eds.)

doi:10.3233/978-1-61499-874-7-308

Applications and Security Risks of Artificial Intelligence for Cyber Security in Digital Environment

Paola AURUCCI[a1]
[a] *University of Eastern Piedmont Amedeo Avogadro, Department of Law, Novara*
paola.aurucci@uniupo.it

Abstract. Everyday more and more devices communicate information over the Internet and the growing demand for protection has become a real challenge for civilization. While security IT systems based on conventional intrusion detection technique are simply not effective in detecting, assessing and countering cyber-threats, the use of AI based systems, thanks to their autonomy, fast paced threat analysis and decision-making capabilities, may guarantee confidentiality, integrity, and availability within the digital environment. However, increased use of AI in cyber defense may create new risks. The aim of this paper is to show experimented applications of AI for cybersecurity, raise awareness on emerging security risks that may hamper the potential of these applications in digital environment and identify possible technological solutions, best practices and legislative interventions to prevent these risks and mitigate intentional and unintentional harmful outcomes of AI based technologies.

Keywords. Artificial intelligence, cybersecurity, security risks, deep learning, artificial neural network, digital environment, malicious use, autonomous systems, intrusion detections systems, lack of control, liability, safety, accountability

1. Introduction

On August 2016, during the DEFCON conference[2], the Paris Hotel in Las Vegas hosted the final round of the Cyber Grand Challenge which was run by the US Defence Advanced Research Projects Agency (DARPA).[3] Seven teams built fully automated artificial intelligence systems to compete in a "no-human allowed" game of "capture the flag"; a fast bug hunting contest on binary code in a highly competitive environment. DARPA's aim was to stimulate development of autonomy in cyber and create unsupervised, autonomous AI hacker, able to quickly discover, prove and resolve bugs in a computer security system. The winning team – US security firm ForAllSecure – received 2 million USD as prize money to continue developing its technology. During the same time period in Las Vegas, the Black Hat conference[4] was held and the security firm SparkCognition unveiled what is said to be the first artificial intelligence powered "cognitive" antivirus

[1] Scholarship holder, Ph.D. University of Milan, Postdoctoral Researcher at the Centre for advanced technology in health and wellbeing (Milan).
[2] DEF CON® 25 Hacking. Website: https://www.defcon.org/html/defcon-24/dc-24-index.html.
[3] U.S. Defence Advanced Research Projects Agency. Website: https://www.darpa.mil/.
[4] Black Hat USA 2016. Website: http://www.blackhat.com/us-16/.

system called DeepArmor[5]. Both these events show how strongly governmental organizations, private companies and security researchers rely on future development of artificial intelligence technique to ensure protection of the cyber sphere from unauthorized intrusions.

As everyday more and more devices communicate information over the internet, the growing demand of protection has become a real challenge for civilization [1]. Conventional IT security measures, which rely on fixed algorithms, speed, skilled machines, and human expertise, are simply not effective in detecting, assessing and countering cyber-attacks. The implementation of AI techniques creates cyber security tools that utilize flexible learning and that are capable of real-time detection and evaluation in order to nearly instantaneously formulate a solution [2]. Drawing on today's advancements in AI techniques and applications, we can tackle a number of major problems raised in the current cyber security scenario, e.g. the detection and prevention of cyber-attacks. This tremendous opportunity comes, however, with an array of risks that require attention and action from legislators, economists, civil servants, regulators, educators and AI researchers. This paper is organized as follow: Section 2 shed the lights on advantages and technological weaknesses **of** intrusion detection and prevention systems (IDPS) used nowadays to ensure cyber protection. Section 3 explains how AI technique could overcome various vulnerabilities and shortcomings of these conventional cyber protection devices and presents some experimented application of AI techniques to cyber defense. Section 4 analyzes security risks that arise from the developments of AI based cybersecurity technologies to better identify potential technological and legal interventions to ensure that the impact of AI on digital environment is net beneficial.

2. Intrusion Detection and Prevention Systems

In 2011, Cisco IBSG researchers predicted that, in a world population of over 7 billion people, there will be 50 billion devices connected to the Internet by 2020 [3]. The growth of the Internet is directly proportional to the number of cyber threats and of potential victims of cyber-attacks and unauthorized intrusions. In addition, these cyber threats sprung from a variety of profiles that can't be targeted in advance, ranging from bored teenagers experimenting with the Internet to rogue states and terrorists deploying direct cyber-attacks. This is the reason why protection of sensitive data from computer intrusions – heather unauthorized access (external intrusions) or malicious use of data (internal intrusion) – today has been regarded as a challenge for civilization [1].

Cyber threats are on the raise and cyber-attacks are becoming everyday more complex thank to the use of multiple redundant attack vectors, to multiply the effects, but also making it more difficult for the response teams to analyse [4]. In order to secure critical business information and safeguard data from increasingly sophisticated and targeted threats, single individuals, governmental organizations and private companies spend millions of Euros in wide variety of technological tools, which help system security administrators protect IT assets. Traditional tools of cyber defence are: firewall, intrusion detection systems (IDS), and intrusion prevention systems (IPS) [5]. While an IDS is designed to identify attacks and alert the system administrator to any malicious event to investigate, an IPS is able to prevent malicious acts or block suspicious traffic on the network. IDS and IPS are not mutually exclusive and for decades have been used

[5] SparkCognition, DeepArmor. Website: https://sparkcognition.com/deeparmor-2/.

concurrently, at least until security experts and vendors realized that these tools could be combined to form an Intrusion Detection and Prevention System (IDPS) [6] capable of ensuring twice the protection [7]. The introduction of IDPS was a significant milestone in the development of effective and practical detection-based information security systems. It is the emblem of good security because it combines monitoring, detection and response and effectively help to achieve security goals of confidentiality, data integrity, authentication and non-repudiations [7]. To get straight to their technical aspects, there are software based IDPS, which are installed on a host computer to analyse and monitor all traffic activities in the system application (Host-Based IDPS) [8] and hardware based IDPS, which are located on an entire network to capture and analyse the stream of data packet sent to a network (Network-Based IDPS) [8]. They are primarily focus on a) detecting and identifying possible intrusions, b) analysing information about the intrusions, c) and attempting to stop the intrusions and report them to security experts/administrators.

2.1. Detection Methods: Anomaly Detection and Signature Detection

The purpose of IDPS is to monitor network traffic for intrusions. These intrusions are recognized through two main detection methods: checking variations in routine behavioural patterns (anomaly detection) or patterns matching (misuse or signature detection) [8].

IDPS anomaly-based detection identifies activities that are different from the reference baseline of accepted network behaviour – given by a human expert – or pattern of normal system activity, learned by the system's analysis of the past activity of the monitored network. Deviations from this baseline cause an alarm to be trigged. On the other hand, IDPS signature-based detection compares potential malicious activity to those that match a defined reference pattern of known attacks or known abnormal behaviour. This process relies on the fact that each intrusion leaves a footprint behind – called signatures – that can be used to identify and prevent the same attack in the future. The human administrator has to create a database of previous attack signatures and known system vulnerabilities that can be used to identify and prevent the same attacks in the future [10]. Usually IDPS combine these two detection methods because of their complementary nature. However, even if these methods are used together, currently used cyber security system aren't able to fulfil the desired characteristic for effectively protecting individuals, organizations and companies from an ever-increasing number of sophisticated attacks.

2.2. Anomaly and Signature Based IDPS, Advantages and Disadvantages

An IDPS should have certain characteristic in order to be able to provide effective and efficient security against serious attacks.

They should be able to: a) guarantee a real-time intrusion detection; b) minimize false positive/negative alarms; c) minimize human supervision d) do constant self-tuning; e) adapt to system changes and users behaviour over time. However, currently used cyber security system aren't able to fulfil this desired characteristic. The most critical and obvious technological "lacks" are: lack of automations [11,12], lack of effective detection [10], lack of predictability of the attack and of effective detection of multiple attacks [13] and lack of flexibility [14]. In sum, combination of speed and skilled physical devices and human expertise intervention is no longer sufficient in defending cyber

infrastructures from more sophisticated cyber threats. In this complex cyber scenario, cyber defence system need to be: a) autonomous; b) able to effectively detect a wide variety of threats without trigger false alarms and reducing the number of false positive/negative rate; c) flexible; and d) robust. Employment of artificial intelligence techniques in cyber security systems can overcome the weaknesses of the commonly used intrusion detection techniques and, as a consequence, could AI play an effective role in the improvement inconsistencies and inadequacies of currently used cyber security systems. In the following section, will be analysed how application of AI techniques can facilitate cyber security measures, especially in terms of effective detection and decreased false positive and false negative rates, the major issues of intrusion management.

3. Artificial Intelligence: The Future Trend in Cyber Security

As seen in Section 2, considering the complexity of the digital environment, IDPS based on conventional intrusion detection technique (like statistical analysis [9] or rule-based) which rely on fixed algorithms, cannot guarantee enough protection for a cyber infrastructure. Their need for a known data pattern for decision making, and continuous human intervention make these cyber security systems ineffective for contrasting dynamically evolving cyber intrusions. All the major issues of security measures analysed in the Section 2 can be overcome by applying AI techniques. AI is a research discipline of computer science that relies on both software and hardware development, that provides method for solving complex problem that cannot be solved without applying some intelligence [8].

Intelligence is simply the capacity to express an appropriate behaviour in response to changes and opportunities in a defined environment [9]. It can be divided into stages of the independent decision-making process as perception, reasoning, and action. Going back to cyber security AI developed flexible techniques which provide learning capabilities and automatic adaptability to conventional systems – hardware and software – used for fighting cyber intrusions. Intelligent cyber security systems can handle and analyse a large amount of information (perception) and, in case of detection of malicious activity, can analyse this information relying on their experience of previous episodes of intrusions (reasoning) and make intelligent decision on which is the proper counteraction (action). All this in real time and without interaction with human analyst-experts. AI researchers have developed a myriad of tools to secure human behaviour and some are already been experimented in the field of cyber security. This section will focus on the potentials and functionalities of the most promising AI tools: artificial neural network-based intrusion prevention and detection systems.

3.1. Artificial Neural Network Based Intrusion Prevention and Detection Systems

ANN is an information processing model that that simulate the structure and the functions of the biological neural system [12]. Like the brain, which is composed of neurons that transmit signal to each other through synapses, via a complex chemical process, the ANN is a net of nodes (processing elements) interconnected by links that transfer numeric data and can transform a set of inputs in a set of desired outputs [13].

If integrated in IDPS for monitoring network traffic, ANN can overcome the shortcomings of other analysed intrusion detection techniques. Thanks to their *inherent speed, their flexibility* and, most of all, their *learning capabilities*, they are able to stop multiple

attackers, quickly predict known pattern of intrusion – even if they do not match the exact characteristic of those that system has been trained to recognize – and reason on information and learned previous episodes of intrusions (experience), to identify new type of attacks pattern, without generate false positive/negative. In few words, IDPSs that rely on ANN are robust, flexible, adaptable and can accurately identify unknown attack without the rule or interaction with the human expert [14]. The results of tests conducted on a neural network offers a promising future in the identification of attack against computer systems. However, security risks triggered by use of AI systems for cyber defense are a matter of concern for both security and legal experts.

4. Security Risks of Artificial Intelligence in Cyber Security

As seen in the introduction and Section 3, more recently AI has been increasingly playing a leading role in cybersecurity's industry.

Since most network-centric cyber-attacks are carried out by highly skilled professionals, which use malware, DDoS, phishing, ransomware, and quickly adopt emerging technologies (e.g. Bitcoins for ransomware payments), private entities and governments are investing in fundamental research to expand the scope of capabilities of AI. Thanks to its autonomy, fast paced threat analysis and decision-making capabilities, AI can enable systems to efficiently detect, defend, and, finally, respond to cyberattack by exploiting the vulnerabilities of antagonist systems. Public curiosity and distress about AI has focused, in particular, on deep, multi-layer machine learning approaches, like neural networks (Section 3.1), seen, nowadays, as essential tools for providing protection. However, less attention has been paid on increasingly pressing security dangers arising from development of AI in cyber security. Next section analyses in details the two major risks that may hamper AI's potential for cybersecurity: possible malicious use [15] of AI systems in digital domains and lack of control [17]. On this basis, interventions are proposed to better investigate, prevent, and mitigate these potential risks. Mapping these criticalities is also vital in order to better appreciate the unique normative challenges of these complex technologies and their impact on current legal systems [18]. These legal issues will be briefly mentions in the Conclusions but are beyond the scope of this paper. They would need further study and a paper on their own to be properly addressed.

4.1. Possible Malicious Use of AI Systems in Digital Domains

A central concern at the nexus of AI and cybersecurity is the potential for malicious uses of AI based systems capabilities.

Because of its generative nature, intelligent systems and the knowledge of how to design them, can be employed for both beneficial and harmful ends. Focusing on the digital security domain, a relevant example is given by AI systems that examine software for vulnerabilities, that might have both positive and malicious applications (e.g. through cyber criminals training systems to hack). Powerful technology falling into the wrong hands (e.g. rogue states, criminal groups and terrorists) would pose grave threats to the security of digital environment. As seen in Section 3, using AI on the defensive side of cybersecurity, makes certain forms of defence more effective and scalable, such as spam and malware detection. But at the same time many malicious actors have natural incentives – which include a premium on speed, labour costs, and difficulties in attracting and retaining skilled labour – to experiment with this powerful technology and develop more

sophisticated AI hacking tools, able to evade detection and creatively respond to changes in the target's behaviour. According to a recent report on the potential malicious use of AI [17] – released by the University of Cambridge and jointly written by twenty-six security experts – further progress and diffusion of efficient AI and machine learning based systems in cyber environment might, first of all, expand the set of actors who are capable of carrying out an attack, the rate at which these actors can carry it out and the set of plausible targets[6]. This claim follows from the qualities of efficiency, scalability, and ease of diffusion that characterize AI based systems[7] and implicate an expansion of existing threats associated with labour-intensive cyberattacks, such as spear phishing[8]. Furthermore, the authors of this report expect that progress in AI will enable new varieties of attacks such as: automated hacking, speech synthesis used to impersonate targets and finely-targeted spam emails using information scraped from social media. This analysis, so far, suggests that the digital environment will change both through expansion of some existing threats and the emergence of new threats that do not exist yet. But report's authors also expect that the typical character of attacks will shift in a few distinct ways. In particular, they think that attacks supported and enabled by progress in AI will be especially effective, finely targeted, difficult to attribute, and exploitative of human vulnerabilities (e.g. through the use of speech synthesis for impersonation), existing software vulnerabilities (e.g. through automated hacking), or the vulnerabilities of AI systems (e.g. through adversarial examples and data poisoning)[9]. Possible changes to the nature and severity of attacks resulting from increasing use of AI will necessitate more vigorous counteroperations [17]. However, this may bring an "escalation" in intensity of attacks and responses, which, in turn, may threaten key infrastructures of our societies. The solution may be to strengthen deterring strategies [19] and discourage opponents before they attack, rather than mitigating the consequences of successful attacks afterward. Yet, necessary (though not sufficient) condition of successfully deterring and punishing attackers is the ability to attribute the source of an attack, a notoriously difficult problem[10][20]. However, the report also identifies a wide range of potential interventions to reduce risks posed by malicious use of AI based systems in digital environment, like: a) developing improved technical measures for formally verifying the robustness end detect most serious vulnerability of the system of the system (e.g through an extensive use of red teaming to discover and fix vulnerability) [21] b) formal verification [22], c) responsible disclosure of development that could be misused (e.g. through extensive use

[6] The use of AI to automate tasks involved in carrying out cyberattacks will alleviate the existing trade-off between the scale and efficacy of attacks [16].

[7] In particular, the diffusion of efficient AI systems can increase the number of actors who can afford to carry out particular attacks. If the relevant AI systems are also scalable, then even actors who already possess the resources to carry out these attacks may gain the ability to carry them out at a much higher rate. Finally, as a result of these two developments, it may become worthwhile to attack targets that it otherwise would not make sense to attack from the standpoint of prioritization or cost-benefit analysis [16].

[8] A phishing attack is an attempt to extract information or initiate action from a target by fooling them with a superficially trustworthy facade. A spear phishing attack involves collecting and using information specifically relevant to the target (e.g. name, gender, institutional affiliation, topics of interest, etc.), which allows the facade to be customized to make it look more relevant or trustworthy [16].

[9] Today's AI systems suffer from a number of novel unresolved vulnerabilities. These include data poisoning attacks (introducing training data that causes a learning system to make mistakes), adversarial examples (inputs designed to be misclassified by machine learning systems), and the exploitation of flaws in the design of autonomous systems' goals.

[10] An example of this problem is given by the failure of the United Nations Cybersecurity Group of Governmental Experts to make progress on norms for hacking in international law. (Korzak, E., UN GGE on Cybersecurity: The End of an Era?, (2017). Available in: https://thediplomat.com/2017/07/un-gge-on-cybersecurity-have-china-and-russia-just-madecyberspace-less-safe/).

of different openness models like pre-publication risk assessment in technical areas of special concern, central access licensing models and regimes that favour safety and security) [23] c) responsible disclosure of AI vulnerabilities; d) envisioning tools to test and improve the security of AI components and use of secure hardware [16]; e) promoting a culture of responsibility through education and ethical statements and standards [17]; f) monitoring of AI-relevant resources [16]. The report also points out a number of research areas where further analysis could develop and refine potential interventions to reduce risks posed by AI malicious use like include privacy protection, coordinated use of AI for public-good security, monitoring of AI-relevant resources, and other legislative and regulatory responses.

4.2. Lack of Control

Further development of AI based defence technologies will increase the complexity of tasks they can perform autonomously, while reducing human ability to understand, predict or control how they operate.

For this reason, the activity of these autonomous systems, to which increasing difficult tasks are delegated, should remain, at least partly, subject to human supervision, either "in the loop" for monitoring purposes or "post-loop" for redressing errors or harms that arise. Progressively less effective control on AI based system used for cyber-defence will increase the risk of unforeseen consequences and errors. This safety challenge has brought a group of security researchers from MIT's Computer Science and AI Laboratory (CSAIL) and a machine-learning start-up known as PatternEx to focus not only on machine automation but also on a better human-computer interaction. In 2016 they designed a neural network based cyber security system, with a human-facing interface that only bothered its human teacher at the right time, called "AI Squared" [25]. It is not a fully automated system, but rather, relies on human control while still being efficient at predicting, detecting and stopping 85% of cyber-attacks with high accuracy, by reviewing data from more than 3.6 billion lines of log files each day. The system first scans the content with unsupervised recurrent neural network techniques and parses data generated by users for potentially odd activity. This process is called "unsupervised learning." Once the neural network has identified the anomalies, it presents its findings to human analysts. The human analyst then identifies which events are actual cyber-attacks and which are not. This feedback is then incorporated into the machine learning system of "AI Squared" and is used the next day for analyzing new logs. This system does not overwhelm the human analysts, and instead, carefully limits the information. The analysts can also give feedback anywhere at any time, either on their smartphones or computers, so that the system can always be learning. In few words, human analysts in AI Squared have the final say back and can control errors and unexpected behavior of the AI system. It is a great example of how even the most advanced AI still needs humans to truly learn—and as a result, still needs designers to craft the language that the human/machine team uses to talk to each other.

Conclusions

This paper has offered a concise analysis on how AI techniques could overcome various vulnerabilities of conventional cyber protection systems (Section 2). Some experimented applications of these techniques to cyber defense are also proposed in Section 3.

Increased use of AI for cyber defense, however, introduces new security risks that may hamper AI's potential for cybersecurity and are a matter of concern for both security and legal experts. Section 4 dwelt on these criticalities that arise from the developments of AI based cybersecurity technologies, in order to shed light on the wide range of potential interventions that can be carried out (Section 4.1) – or that are already developed (Section 4.2) – by AI researchers and practitioners so as to tackle these risks. Yet, these technological solutions to be effective must be supported by a clear legislative and regulatory framework, able to reduce threats triggered by these new technologies and increase stability without hindering research and development in the field [26]. To prevent security risks, regulators and policy makers should learn from other domains with longer experience [27] and put in place rules ensuring safety of products in the commercialisation phase (e.g. through testing, certification and insurance mechanisms), coupled with well-designed financial incentives and liability safeguards to mitigate intentional or unintentional harmful outcome of AI applications. In the cyber security domain, policy and regulations may also mitigate the dangers of lack of control on AI systems by ensuring proportionality of responses, the legitimacy of targets, and a higher degree of responsible behavior.

Contrary to popular belief, the AI industry development does not take place in a regulatory vacuum and a *de facto* AI legal framework already exists [28]. General Product Safety Directive 2001/95/EC (GPSD)[11] and the Product Liability Directive 85/374/EEC[12] apply, for example also to innovative businesses working with AI. However, the most salient characteristics of AI technology, like their unpredictable behaviour or the complexity of the ecosystem behind machine learning [18], trigger new legal issues that make the current EU regulatory framework particularly unsuited to address risks brought about by the use of intelligent and autonomous systems[13]. Admittedly, these new legal challenges of AI systems vary in accordance with the field under examination: international law, criminal law [29], civil law, both contract and tort law, administrative law and so forth. Focusing on civil law, AI security systems may rise legal issues that include, but are not limited, to liability and data governance. With regard of liability, scholars have stressed time and again the complexity of distributed responsibility [30], drawbacks of strict liability policy [18] – that may hinder technological research – and the need of new methods of accountability and insurance policy. A concise analysis of legal issues goes beyond the scope of this paper, whose primary aim was to draw attention to the security risks triggered by the use of AI in cyber defence, rather than specific policy proposals. In this respect, it is interesting the recent call for application concerning the Expert Group on Liability and New Technology recently published by the EU commission[14]. Still, it is crucial to start shaping policy and regulations for the use of AI in cyber environment while this technology is nascent. To do so, close collaboration between legislators and technical researchers and mechanisms of legal flexibility [26] are vital to shape regulation able to prevent and mitigate potential AI risks avoiding the implementation of measures that may hamper research progress.

[11] Directive 2001/95/EC on general product safety. Available at: http://ec.europa.eu/consumers/consumers_safety/product_safety_legislation/index_en.htm.

[12] Directive 85/374/EEC on liability for defective products. Available in: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31985L0374.

[13] For this reason, the EU Commission started the process for the amendment of the aforementioned directive in September 2016 [18].

[14] EU Commision: http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=615947.

References

[1] S. Dilek, H. Çakir, M. Aydin, *Applications of Artificial Intelligence Techniques to Combating Cyber Crimes: A Review. In: International Journal of Artificial Intelligence & Applications (IJAIA),* Vol. 6, No. **1** (2015). 21–39. DOI: 10.5121/ijaia.2015.6102.

[2] I. Mukhopadhyay, M. Chakraborty, *Hardware Realization of Artificial Neural Network Based Intrusion Detection & Prevention System. In: Journal of Information Security*. 2014. 154–165. DOI: 10.4236/jis.2014.54015.

[3] D. Evans, *The internet of things. How the next evolution of the internet is changing everything*. Available at: https://www.cisco.com/c/dam/en_us/about/ac79/docs.pdf.

[4] T. Yadav, A.M. Rao, *Technical Aspects of Cyber Kill Chain*. In: Abawajy J., Mukherjea S., Thampi S., Ruiz-Martínez A. (eds), *Security in Computing and Communications. SSCC 2015. Communications in Computer and Information Science, Springer, Cham*, Vol. **536 (**2015): 438–452. DOI: 10.1007/978-3-319-22915-7_40.

[5] A. Fuchsberger, *Intrusion Detection Systems and Intrusion Prevention Systems*. In: *Information Security Technical Report. Elsevier, Amsterdam*, Vol. **10** (2005), 134–139. DOI: 10.1016/j.istr.2005.08.001.

[6] K.A. Scarfone, P.M. Mell, *Guide to Intrusion Detection and Prevention Systems (IDPS). In: National institute of standards and technology special publication* (2007). 800–94. DOI: http://dx.doi.org/10.6028/NIST.SP.800-94.

[7] F. Farokhmanesh, Intrusion Detection and Prevention Systems (IDPS) and Security Issues. In: IJCSNS International Journal of Computer Science and Network Security, Vol. 14, n. **11** (2014), 80–84.

[8] B. Santos Kumar et al., *Intrusion Detection System – Types and Prevention. In: (IJCSIT) International Journal of Computer Science and Information Technologies*, Vol. 4, n. **1** (2013). 77–82. Available at: http://ijcsit.com/docs/Volume%204/Vol4Issue1/ijcsit2013040119.pdf.

[9] C. Wang et al., Statistical technique for online anomaly detection in data canters. In: *IFIP/IEEE International symposium on integrated network management*, 2011. Available at: http://www.hpl.hp.com/techreports/2011/HPL-2011-8.pdf.

[10] A.S. Ashoor, S. Gore, Difference between Intrusion Detection System (IDS) and Intrusion Prevention System (IPS). In: Wyld D.C., Wozniak M., Chaki N., Meghanathan N., Nagamalai D. (eds), Advances in Network Security and Applications. CNSA 2011. Communications in Computer and Information Science, Springer, Berlin, Heidelberg, Vol. **196** (2011). 497–501. DOI**:** https://doi.org/10.1007/978-3-642-22540-6_48.

[11] H. Shrobe et al., *New Solutions for Cybersecurity*, MIT press Cambridge, 2017.

[12] J. Frank, Articial Intelligence and Intrusion Detection Current and Future Directions, 1994. Available at: http://home.eng.iastate.edu/~guan/course/backup/CprE-592-YG-Fall-2002/paper/intrusion/ai-id.pdf.

[13] K.S. Devikrishna, B.B. Ramakrishna, International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 3, Issue 4, Jul–Aug 2013, 1959–1964.

[14] M.H. Bhuyan, D.K. Bhattacharyya, J.K. Kalita, *Anomaly based Intrusion Detection Using Incremental Approach: A Survey, in Network Security: Issues, Challenges and Techniques, Narosa Publishing House, India*, 2010. 112–125. Available at: http://www.cs.uccs.edu/~jkalita/papers/2013/BhuyanMonowarIEEECOMST.pdf (Last accessed: August 15, 2017).

[15] M. Brundag et al., *The malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation*, Cambridge University Press, 2018. Available at: arXiv:1802.07228v1.

[16] G.Z. Yang et al., *The Grand Challenges of Science Robotics, Sci. Robot*, **3** (2018), eaar7650. DOI: 10.1126/scirobotics.aar7650.

[17] U. Pagallo, *From Automation to Autonomous Systems: A Legal Phenomenology with Problems of Accountability*. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Invited Speakers. 17–23. DOI: https://doi.org/10.24963/ijcai.2017/3.

[18] M.R. Taddeo, "The limits of deterrence theory in cyberspace". *Philos. Technol*. **10** (2017) 1–17.

[19] E. Korzak, "UN GGE on Cybersecurity: The End of an Era?", 2017. Available in: https://thediplomat.com/2017/07/un-gge-on-cybersecurity-have-china-and-russia-just-madecyberspace-less-safe/).

[20] H. Abbass et al., *Computational Red Teaming: Past, Present and Future, IEEE Computational Intelligence Magazine*, Volume 6, Issue 1, 2 (2011), 30–42.

[21] C. Baier, Katoen, *Principles of Model Checking*. *Cambridge: MIT Press*. J. (2008).

[22] P. Eckersley, Y. Nasser et al., *Help EFF Track the Progress of AI and Machine Learning, Electronic Frontier Foundation* (2017). Available at. https://www.eff.org/deeplinks/2017/06/help-efftrack-progress-ai-and-machine-learning.

[23] U. Pagallo, M. Durante, *The Pros and Cons of Legal Automation and its Governance, European Journal of Risk Regulation, Cambridge University Press*, Volume 7, Issue 2, 1 (2017). 322–334. DOI: https://doi.org/10.1017/S1867299X00005742.

[24] K. Veeramachaneni et al., "AI2: Training a big data machine to defend". Available in: https://people.csail.mit.edu/kalyan/AI2_Paper.pdf.
[25] U. Pagallo, *Even Angels Need the Rules: AI, Roboethics, and the Law". In Gal A Kaminka et al (eds), ECAI (2016). Frontiers in Artificial Intelligence and Applications*, 209–215. IOS Press, Amsterdam 2016. DOI: 10.3233/978-1-61499-672-9-209.
[26] U. Pagallo, *The laws of robots: crimes, contracts, and torts. Springer Science & Business Media*, 2013.
[27] M. Brundage, J. Bryson, *Smart polices for artificial intelligence*. Available in: https://arxiv.org/abs/1608.08196.
[28] P.M. Freita, F. Andrade, P. Novais, Criminal Liability of Autonomous Agents: From the Unthinkable to the Plausible. In: Casanovas P., Pagallo U., Palmirani, M., Sartor G. (eds), AI Approaches to the Complexity of Legal Systems. Lecture Notes in Computer Science, Vol. **8929** (2014). Springer, Berlin, Heidelberg.
[29] C. Karnow, "Liability for Distributed Artificial Intelligence", Berkeley Technology and Law Journal, 11 (1996). 147–183.
[30] D.J. Gunkel, The Machine Question, critical perspectives on AI, robots, and ethics. MIT Press Cambridge 2017.