TALLINN UNIVERSITY OF TECHNOLOGY

School of Business and Governance

Faculty of Social Sciences

Valeria Gaiduk

LEGAL RESPONSIBILITY IN THE CONTEXT OF AI-WRITTEN MALWARE

Master's Thesis

Supervisor: Agnes Kasper, Ph.D.

Tallinn 2020

I hereby declare that I have compiled the thesis/paper independently

and all works, important standpoints and data by other authors

have been properly referenced and the same paper

has not been previously presented for grading.

The document length is .....18045...... words from the introduction to the end of the conclusion.

Valeria Gaiduk …...............................

(signature, date)

Student code: 162914HAJM
Student e-mail address: v.gaiduk@hotmail.com

Supervisor: Agnes Kasper, PhD:
The paper conforms to requirements in force

....

(signature, date)

Chairman of the Defence Committee:

....

(name, signature, date)

TABLE OF CONTENTS

ABSTRACT

The aim of this paper is to examine and to determine the limits of criminal and civil responsibility for malware created by AI between the creator and trainer of AI model in different scenarios, including intentional creation, accidental foreseeable, and accidental unforeseeable creation. It is necessary to understand the difference between the obligations of the creator and trainer of AI-created malware in order to assist states in fulfilling their duties.

There is a liability problem for AI-written malware in criminal and civil law. Many actors may be involved in the creation and distribution of malware. The transnational use of AI in the modern world emphasizes that malware can become a real threat. Therefore, it is necessary to correctly identify the participants who worked on the creation and distribution of malware, to limit possible damages as well as to prosecute the perpetrators.

The following research will include the analysis of scientific literature and legal acts based on several authors. This research is based on primary and secondary sources. The existence of intent and negligence will be factored into criminal and civil liability analysis. After the analysis, suggestions will be given for establishing a legal system of liability for the AI-perpetrated malignancies.

Given the lack of regulations at the national level related to the prosecution of the perpetrators who instruct AI to create malware, the Member-States need to adhere to the same constituent elements of criminal and civil offenses based on precedents related to cyber-attacks. National laws must review various circumstances during the legal process. Therefore, the author believes that it would be reasonable to improve the national legislation and the Directive 2013/40/EU, since there are many sanctions and penalties for legal entities, but not for individuals and states as well as it would be useful to use the theory of "deep pocket" and update the General Product Safety Directive 2001/95/EC and the Product Liability Directive 85/374/EEC to ensure a software manufacturer’s liability.

AI-created malware, AI, malicious code, criminal responsibility, civil responsibility, prosecution.

INTRODUCTION

Artificial intelligence (AI) is increasingly important in modern life. It has a variety of actual and potential uses, ranging from traffic light systems to surgery. The applications for using AI are virtually limitless, but its increased use can bring vulnerabilities. There are many dangers associated with the growing use of AI, such as the susceptibility of vital systems to outside attacks, loss of sensitive data, and some have even suggested the possibility of world dominance by AI, notably Stephen Hawking.

Aside from theories of future human-robot warfare, the use of AI can cause mayhem even today, notably when its use is accompanied by malware. Most people with access to computers, smartphones, and the internet have heard of the term malware and some of its subspecies such as computer viruses and trojans. Malware, or malicious computer programs are responsible for ever-increasing losses of revenue, trade secrets, state secrets, identity thefts, leakage of passwords and codes, uncovering of sensitive personal data, alteration and manipulation of data, forgery, theft of funds and many more. The increasing use of AI in everyday systems can cause even more immediate consequences if such manipulations occur, for example, traffic accidents or medical equipment, failures can become more pronounced with increasing automation.

Most malware is written by human coders who can limit the number of potential threats; however, the automation of many tasks brought by the development of AI can potentially skyrocket malware production. Computer programs are composed of sequenced commands aimed at causing the programmable machine to perform some activity. Coders typically use high-level programming languages for writing software programs as they have a similar syntax to human language and are therefore easier to understand and use. Programming languages have apparent rules, and logic-based systems such as AI are perfect for operating on logic-based tasks, making them ideal for program-writing. Many coders, however, have their specific „handwriting“ for creating new software, since the particular sequence of commands can differ, and there can be various creative ways how to write a program. While a standard writing software can sequence commands based on rules written inside the writing program, AI-based systems can learn the patterns of human coders and develop new ways of writing programs. This can potentially cause the diversification of malware

There can be several ways how AI begins creating malware. First, AI can be created or modified by someone with malicious intent who purposefully creates a system for manufacturing malware. Malware can also be created unintentionally either by mistakes in coding non-malicious programs or by AI system acting spontaneously.

The main idea of this research is to examine and to determine the limits of criminal and civil liability for malware created by AI between the creator and trainer of AI model in different scenarios, including intentional creation, accidental foreseeable and accidental unforeseeable creation. Thus, understanding the nature and scope of obligations to malware created by AI between the creator and trainer of AI model is an important step in helping states to meet their duties.

The research problem states that there is a liability problem for AI-written malware in criminal and civil law. Many actors can be involved in the creation and spread of malware and the increasing use of AI in the modern world predicts that malware will become an increasing threat, becoming more diverse and widespread and penetrating to areas previously safe. For this reason, it is important to identify different actors in different situations of malware creation and spread. This way, it can be possible to limit the damages caused by malware and sanction those responsible.

The relevance of the following research AI is increasingly important in modern life. It has a variety of actual and potential uses, ranging from traffic light systems to surgery. The applications for using AI are virtually limitless, but its increased use can bring vulnerabilities. There are many dangers associated with increasing use of AI, such as the susceptibility of vital systems to outside attacks, loss of sensitive data, and some have even suggested the possibility of world dominance by AI, notably Stephen Hawking.

The main research question of the following thesis is: Which actors would bear liability for damages caused by malicious code generated by AI?

The following research will include the analysis of scientific literature and legal acts based on several authors. This research is based on primary and secondary sources. The primary sources include official documents released by the international organizations, legislation and policy documents, as well as case law will be used to determine the different roles that various actors can have in AI-mediated malicious code creation and distribution. The secondary sources cover books and legal journals that are carefully selected with the criteria of their reliability. Afterwards, the existence of intent and negligence will be factored into criminal and civil liability analysis. Tort rules will be discussed. After the analysis, suggestions will be given for establishing a legal system of liability for AI-perpetrated malignancies.

The thesis will cover several aspects of criminal and civil responsibility for malware created by AI between the creator and trainer of the AI model in different scenarios and will be divided into three parts. Firstly, the explanation of the basics of computer programming and the nature of software and programming languages. Also, principles of machine learning, possibilities and limitations, creators of AI systems, examples of applications using AI, future possibilities will be examined. Secondly, current liability for AI-created malware will be reviewed, as well as liabilities of programmer, trainer, operator, and liability of oversight if AI creates malware on its own. Thirdly, the proposal for a liability system for AI-based malware will be discussed. In the Conclusion section, the possible solutions will be discussed.

1. SOFTWARE

Software is a system of instructions for a computer that make the computer produce some result.[1]

1.1 Artificial Intelligence

Artificial intelligence (AI) is a subfield of computer science that has the purpose of enhancing human intelligence capabilities, mostly thanks to deep learning and high-performance computing.[2] AI is any created software system that simulates human thinking on a computer or other devices such as home management systems integrated into household appliances; robots; autonomous cars; and unmanned aerial vehicles. [3] However, Stuart J. Russell and Peter Norvig identify AI as a system that is able to think, can act where necessary, and be rationalized as a human.[4]

AI has been used in several fields, mostly for making models of processes, making human analysis better and increasing automation. AI-based systems can be divided into two parts: the first can be based only on software, operating in the virtual world such as voice assistants, image analysis software, search engines, speech and face recognition systems, while other AI-based systems can be embedded in hardware devices such as complicated robots, autonomous machines and IoT applications.[5]

Robotics is one of the key areas of AI use, also modern facial and speech recognition but also anti-malware functions in cybersecurity.[6] In forensics, AI is tested for finding out the camera models

[1] Kirby, C. A., Defining Abusive Software to protect Computer Users from the Threat of Spyware. Computer L. Rev. & Tech. J., 10, 2006, p 309. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/comlrtj10&i=287 , 25 March 2020

[2] Li, J.-H. Cyber security meets artificial intelligence: a survey. Front. Inform. Technol. Electron. Eng., 19(12), 2018, p 1462.

[3] McCarthy, J. What is artificial intelligence? Stanford University, Computer Science Department. 2007, p15. In: Čerka, P., Grigienė, J., Sirbikytė, G. (2015). Liability for damages caused by artificial intelligence. Computer Law & Security Review: The International Journal of Technology Law and Practice, 31(3), p378

[4] Sherman, C. R. The Surge of Artificial Intelligence: Time To Re-examine Ourselves. Implications: Why Create Artificial Intelligence? 1998. In: Čerka, P., Grigienė, J., Sirbikytė, G. Liability for damages caused by artificial intelligence. Computer Law & Security Review: The International Journal of Technology Law and Practice, 31(3), 2015, p377

[5] Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe, Brussels, 25.4.2018 COM(2018) 237 final.

[6] Li (2018), *supra nota* 2, 1463.

where images come from.[7] Language processing is another field where AI is beneficial, and there is more and more AI use in law too.[8] But some think that there are many tasks in law that AI cannot take over.[9] In healthcare, AI is used in some places to identify potential healthcare fraud, to detect patterns and to learn about clinical trial data.[10] Machine learning is an integral part of AI development, but there are limitations that are especially important in cybersecurity.[11]

There are limitations that affect AI systems and specialists that are associated with knowledge specific to this problem. The relevant expertise is needed to improve the system and data. In this regard, people and AI systems must know the latest knowledge in order to determine which parts of their previous experience is outdated. Knowledge in the field of robotics and the automotive industry does not change rapidly, but in the world of cybersecurity, knowledge of exploits and fixes is changing daily.[12]

The algorithms that are used in machine learning must have the target features defined, and the success of the machine learning depends on how well the features are recognized.[13] Machine learning has developed into deep learning, relying on deep neural networks where features to discover are not fed to AI, but AI is trained with data.[14] This has improved AI functionality significantly.[15] The advantages of deep learning are its ability to notice new patterns and attacks, accept any file types.[16] AI uses pattern recognition to adapt its algorithms and to improve this recognition even more.[17] In the past, the forgetting of the previous task and the knowledge connected with it was a severe challenge to AI building, but it has been resolved.[18]

[7] Athanasiadou, E., et al., Camera Recognition with Deep Learning. Forensic Sci. Res., 3(3), 2018, pp 210-218. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/forsr3&i=212, 25 March 2020

[8] Semmler, S., Rose, Z., Artificial Intelligence: Application Today and Implications Tomorrow. *Duke L. Rev.*, 16, 2017-2018, pp 85-99. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/dltr16&i=85, 25 March 2020

[9] Wendel, B. W. The Promise and Limitations of Artificial Intelligence in the Practice of Law. *Okla. L. Rev.*, 72, 2019, pp 24-25.

[10] Terry, N. P. Appification, AI, and Healthcare's New Iron Triangle. J. Health Care L. & Pol'y, 20(2), 2018, p133. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/hclwpo20&i=127, 25 March 2020

[11] Li (2018), *supra nota* 2, 1463.

[12] Kingston, J. Artificial Intelligence and Legal Liability. Australian Product Liability Reporter, 28(3), 2018, p276. Retrieved from https://arxiv.org/pdf/1802.07782.pdf, 8 April 2020

[13] Li (2018), *supra nota* 2, 1463.

[14] *Ibid*, 1463

[15] *Ibid*.

[16] *Ibid*.

[17] Davis, J. P., Law without Mind: AI, Ethics, and Jurisprudence. Cal. W. L. Rev., 55(1), 2018, p179. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/cwlr55&i=175, 25 March 2020

[18] Terry (2018), *supra nota* 10, 138.

Modern AI that uses deep learning can function similarly to the networks of neurons in the brain, and that is why it is called a neural network.[19] The learning experience in deep learning technology is formed by many neural parts that are interconnected to each other just like neurons in the brain, and the connections are reformed based on the results of the analysis.[20] When there are many layers of these artificial neurons then the combinations of possibilities are bigger, and AI is more powerful.[21] This is the deep reinforcement learning where the correct results are preferred, and incorrect ones are punished.[22] Learning may be supervised, meaning that there is input data that is explained or labeled or there is unsupervised data that comes raw from the surrounding.[23] It is also possible to use data from computer simulations.[24] A similar approach to neural networks is the genetic approach where parameters are put into gene-sets, and then the cost-benefit analysis is used to correct parameter values, and mutagenesis adds variable parameters to make things more diverse.[25] This genetic approach is used to detect malware using AI.[26]

Based on all of the above definitions, it becomes clear that AI is different comparing to the ordinary computer algorithms. Scientists strive to improve artificial intelligence so that it becomes self-learning and can accumulate personal experience or possess machine learning. In this connection, such a unique function will allow AI to act in the same situations differently based on experience, which brings AI closer to human rational thinking and behavior. Such methods, and in particular cognitive modeling, provide greater flexibility and allow the creation of programs similar to the processes of human brain activity. And although this seems unbelievable, Hayes claims that there are already certain programs that mimic certain processes of the human brain, the functioning of which is based on an artificial neural network.[27]

[19] Scharre, P. Killer Apps: The Real Dangers of an AI Arms Race. Foreign Aff., 98(3), 2019, p 136.
[20] *Ibid*., 136.
[21] *Ibid*.
[22] Weyhofen, C., Scaling the Meta-Mountain: Deep Reinforcement Learning Algorithms and the Computer-Authorship Debate. *UMKC L. Rev.*, Vol. 87, No. 4, 2019, pp 989-990. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/umkc87&i=1013, 25 March 2020
[23] Scharre (2019), *supra nota* 19, 136-137.
[24] *Ibid*, 137.
[25] Schuster, W., M. Artificial Intelligence and Patent Ownership. Wash. & Lee L. Rev., 75(4), 2018, pp 1955-1958. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/waslee75&i=1991, 25 March 2020
[26] Pfeffer, A., et al. Artificial Intelligence Based Malware Analysis. ArXiv. 2017. Retrivered from http://arxiv-export-lb.library.cornell.edu/pdf/1704.08716, 27 Jan 2020.
[27] Čerka, P., Grigienė, J., Sirbikytė, G. Liability for damages caused by artificial intelligence. Computer Law & Security Review: The International Journal of Technology Law and Practice, 31(3), 2015, pp*376-389*

1.2 Malware

The term malware refers to malicious software, a harmful software.[28] Malware has been in the minds of people from the 1988 attacks on universities and research centers in the U.S.[29] The standardization of computing technology in recent history has made computers easier to use but has also made them more vulnerable to malware. Compatibility of systems means that data storage and retrieval are the same or similar and that makes it easier for malware to spread.[30]

Malware is a software type used for malicious activities.[31] Malware can enter a computer system in various ways, like clicking on links that lead to websites downloading malware or opening infected emails. When a system is infected, it is also possible to add backdoors to that system for easy access without user knowledge.[32] Malware often uses the vulnerabilities found in systems and networks to create entry points. Many different vulnerabilities exist such as those who allow to change the computer's memory and those who target and change webpages.[33] Advanced malware is even capable of avoiding detection by special tools.[34]

There are many tools that can detect malware. Scanners read through executable files and search for viral signals that are small unique parts of code that researchers have identified for known viruses.[35] Activity monitors check for behaviors that viruses often do, which makes them useful for catching unknown viruses but only after virus execution has started already.[36] Integrity checkers analyze if a file is modified but this can be inaccurate.[37] Heuristic detectors analyze behavior, look from likely locations and use logic to determine the likelihood of a virus infection of the system.[38]

[28] Rodriguez, M. All Your IP Are Belong to Us: An Analysis of Intellectual Property Rights As Applied to Malware. Tex. A&M L. Rev., 3(3), 2016, p 664. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/twlram2015&i=691, 25 March 2020

[29] Hansen R. L. The Computer Virus Eradication Act of 1989: The War against Computer Crime Continues. Software L.J., 3, 1989-1990, pp 717-718.

[30] *Ibid*., 722-723.

[31] Arivudainambi, D., et al. Malware traffic classification using principal component analysis and artificial neural network for extreme surveillance. Computer Communications, 147, 2019, p 50.

[32] *Ibid*.

[33] De Villiers, M. Free Radicals in Cyberspace – Complex Liability Issues in Information Warfare. Nw. J. Tech. & Intell. Prop., 4(1), 2005, p 25. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/nwteintp4&i=17, 25 March 2020

[34] Arivudainambi (2019), *supra nota* 31, 50.

[35] De Villiers (2005), *supra nota* 33, 21.

[36] *Ibid*., 22.

[37] *Ibid*., 22-23.

[38] *Ibid*., 23.

Different types of malware exist. The list of harmful or potentially harmful programs keeps getting longer. The technological development in the computing industry is swift. It is like an arms race that is going on between hackers-attackers and cybersecurity companies. One of the most commonly known types of malware is computer viruses. They behave quite the same way as biological viruses, meaning that they replicate themselves and invade new systems.[39] They do this by attaching to another computer program located on the target computer, executing with the program it is tied to, copy themselves, and then the copies get connected to other programs in the same and different computers in the network.[40] They sometimes have the side effect of damaging data.[41] Worms are those computer viruses that do not have to stick to a program but can copy themself and spread on its own.[42] Worms can destroy files or slow down the connection because of its spreading behavior.[43] At first, they stick to the executable programs, and when these programs are executed, they copy themself and find the other executable programs to glue themself to.[44] The virus then checks if the conditions of action have been met and if they have, it activates the damage-causing commands.[45] Viruses can enter into programs by overwriting them, adding themselves to the ends of the code, or entering into parts of the program that do not contain values.[46] Execution of virus can cause direct data damage, information theft, no damage at all, or consume electricity and memory.[47] Viruses used to travel through data storage media but nowadays, they mainly travel through the Internet.[48] Many technologies that are used to fight viruses, some specific ones that recognize their particular pattern, and some generic ones that recognize the pattern of action of the virus.[49]

Another commonly known type is Trojan horses. They seem to be a typical software or a file but contains a code that can run other malware types or make a remote connection with other devices.[50] Also, in the context of protection against liability for AI systems, there are cases when a person accused of committing cybercrime successfully put forward a line of defense that his computer

[39] Rodriguez (2016), *supra nota* 28, 667.

[40] De Villiers, (2005), *supra nota* 33, 18.

[41] *Ibid*.

[42] *Ibid*.

[43] *Ibid*.

[44] Kroczynski, R. J. Are the Current Computer Crime Laws Sufficient or Should the Writing of Virus Code Be Prohibited? Fordham Intell. Prop. Media & Ent. L.J., 18(3), 2008, p 824. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/frdipm18&i=825, 25 March 2020

[45] De Villiers (2005), *supra nota* 33, 19.

[46] Kroczynski (2008), *supra nota* 44, 825-826.

[47] De Villiers (2005), *supra nota* 33, 19.

[48] *Ibid*., 20.

[49] *Ibid*.

[50] Rodriguez (2016), *supra nota* 28, 667.

was hacked by a trojan or some other malicious virus that committed crimes without the knowledge of the defendant, by using his computer.[51] In *Regina v Aaron Caffrey*[52] case the defendant claimed that his computer was hijacked by hackers by using a Trojan horse to remotely control his PC,[53] and eventually, a Trojan wiped itself from the computer. The defendant's lawyer successfully convinced the jury that this could happen because no one of them could understand the technical arguments,[54] even though the prosecution argued that no trace of Trojan attack was found.[55]

It is technically possible that the malware was installed but was ultimately wiped and there is no evidence of this. Moreover, even when a file is deleted, the data associated with it will still be stored on the computer until it is manually wiped, and the system is restarted. To wipe such files requires a special application. However, if a hacker removes traces of his attack, he must also delete files associated with the "wiping tool",[56] since such a tool can leave certain signatures and cannot delete itself. If the operating system does not erase the default data and temporary files, then in the swap data you can detect signs of malware and the wiping tools[57].

However, it may happen that the malware data will not be detected, but the traces of wiping will be found. In this case, it is difficult to draw a clear-cut conclusion that the owner of the PC is liable or not for the attack. People working with confidential information and trade secrets use similar tools to wipe confidential data in case of computer theft. Also, a person can engage in illegal activities and eventually will try to hide his activities. In this case, it will be challenging to say whether a person tried to hide something or whether malware was present in the system.[58]

It is evident that when this investigation was conducted, there was not enough knowledge on this topic. Due to the lack of precedents and available information regarding such cases, the accused managed to escape punishment. However, considering this case, it can now be argued that the prosecution had enough evidence to prove their case.[59]

[51] Kingston (2018), supra nota 12, 273.
[52] Southwark Crown Court, Unreported, *Regina v Aaron Caffrey,* 17.10.2003
[53] Brenner, W.S., Carrier B., Henninger J. The Trojan Horse Defense in Cybercrime Cases, Santa Clara High Tech. L.J., 21(1), 2004, p6. Retrieved from: http://digitalcommons.law.scu.edu/chtlj/vol21/iss1/1, 8 May 2020
[54] *Ibid*.
[55] *Ibid*., 6.
[56] *Ibid*., 49.
[57] *Ibid*., 50.
[58] *Ibid*., 50-51.
[59] *Ibid*., 6.

There was another case in the United Kingdom, *R v Green*[60] when there was found a child pornography on Mr. Green's hard drive, but he was not aware of it and did not give his permission to put the child pornography on his PC. During the investigation, the IT experts found 11 Trojan horse programs on Green's PC. Eventually, based on the previous precedent *Regina v Aaron Caffrey* and substantial evidence, the court released the accused. [61]

A type of malware that has recently become more known is spyware. That one collects data from the infected system and sends it to someone else to be used for different purposes.[62] The terms appeared in 1999 to describe programs that sent personal information out.[63] Someone who is browsing webpages could download the spyware unintentionally, not even knowing about it.[64] There are two things that the spyware does. It violates the victim's privacy and it takes away the control that the victim has over the computer.[65] Some consider cookies to belong to spyware too.[66] Another type of malware that has recently become widely known is ransomware. It is usually transmitted via email attachments that execute when the attachment is opened.[67] Ransomware encrypts files on the victim's computer and then displays a message telling victim how to pay for the decryption using bitcoin.[68] Ransomware is even available as a service (RaaS), made by technically skilled programmers and sold to users who use them.[69] Some of those RaaS ransomwares like Satan are very user friendly and even have customer support.[70]

There are malware types that are lesser known to those who are not very literate in computer terminology. One of those is rootkits. They are files and scripts that give their placer the ability to keep accessing the system to get data or do something else.[71] The rootkits are often combined with other malware such as a virus or a Trojan and they can be used to set up a botnet.[72] One of the most known rootkits is ZeuS which is configurable by a user with additional modules that the

[60] Exeter Crown Court, Unreported, *R v Green*, October 2003

[61] Brenner (2004), *supra nota* 53, 7.

[62] Rodriguez (2016), *supra nota* 28, 667.

[63] Kirby (2006), *supra nota* 1, 291.

[64] *Ibid*.

[65] *Ibid*., 292.

[66] Kroczynski (2008), *supra nota* 44, 823.

[67] Harkins, M., Freed, A. M. The Ransomware Assault on the Healthcare Sector. J.L. & Cyber Warfare, 6(2), 2018, p151. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/jlacybrwa6&i=339, 25 March 2020

[68] *Ibid*., 151.

[69] *Ibid*., 152.

[70] *Ibid*., 152-153.

[71] Rodriguez (2016), *supra nota* 28, 667-668.

[72] *Ibid*., 668.

users can buy, those modules can give extra functions like giving backdoor access to infected computer or even taking over the infected computer.[73]

A botnet is a group of computers called drones or zombies that follow the owner's commands through a Command and Control server.[74] They are used for large data processing functions that can be everyday operations but can also be used for bad purposes like DDoS attacks.[75]

The effects of malware can be devastating. One of the most vulnerable sectors is healthcare. Patient medical data is placed into electronic records that provide access to sensitive personal information for healthcare workers and patients. In a standard system that is very widely used, intrusion can have very serious consequences. According to Article 3 of Directive 2013/40/EU "Member States shall take the necessary measures to ensure that, when committed intentionally, the access without right, to the whole or to any part of an information system, is punishable as a criminal offence where committed by infringing a security measure, at least for cases which are not minor".[76]

Harkins and Freed have written that Electronic Health Records have made it easier for attackers to access medical data of patients and then putting it on sale, holding it hostage for money, for identity theft purposes etc.[77] The value of health records is larger than financial information, making healthcare sector a tempting target.[78] Harkins and Freed write that impersonation techniques are able to fool even very smart people like doctors.[79] The recent WannaCry attack caused big disturbances in UK's health system and showed how poor the information security is in the NHS.[80] Despite danger, about half of healthcare centers do not have response plans for those events and about half have experienced data leaks within one year.[81]

[73] *Ibid*., 668-669.
[74] *Ibid*., 668.
[75] *Ibid*.
[76] OJ L 218, 14.8.2013
[77] Harkins (2018), *supra nota* 67, 150.
[78] *Ibid*., 154.
[79] *Ibid*., 155.
[80] *Ibid*., 157-158.
[81] *Ibid*., 161.

1.3 Conclusion of AI and malware

There are many good uses of AI but there can be bad ones too. Many people are afraid that mankind may eventually not be able to control AI.[82] Even now, the bad uses of AI must be analyzed. AI is now used to detect malware from power use patterns of smartphones to discover secret communication channels.[83]

AI can also be used a lot in malware manufacturing. This is one of the abilities of AI – to create a lot of malware and to create malware that develops itself. There are different processes in malware generation that can be automated. One of those is exploit generation.[84] It used to be the case that exploiting bugs was done by a person who had to analyze if a software bug can be used as a weakness.[85] The network that is hacked needs to be scanned for open ports while hiding the attacker's identity.[86] After that, the attackers often place software to collect data that is necessary for making the attack happen.[87] Searching for bugs and then deciding if it is possible to exploit them for access or not took quite some time. Those who wanted to find, and exploit vulnerabilities had to make effort and it might not have worked at all. However, with automation this can be done in a much faster and easier way. Blended attacks that are like Nimda and CodeRed are designed to find new vulnerable points in the computer system.[88] The need to study the technology behind computer systems and network properties to find vulnerabilities does not exist anymore.[89]

Now, those who want to hack into a computer system or network do not have to know much about the technology at all. Tools for making these attacks happen are available in a ready to use form.[90] These are called malware kits or exploit kits.[91] This means that it is not possible to assume anymore that the attackers have quite a lot of knowledge. Anyone could be responsible for an

[82] Martinez, R. Artificial Intelligence: Distinguishing Between Types & Definitions. Nev. L.J., Vol. 19, 2019, p 1027.

[83] Caviglione, L., *et al.* Seeing the Unseen: Revealing Mobile Malware Hidden Communications via Energy Consumption and Artificial Intelligence. *IEEE Transactions on Information Forensics and Security*, Vol. 11, No. 4, 2016, pp 799-810.

[84] Arivudainambi (2019), *supra nota* 31, 50.

[85] Avgerinos, T., et al. AEG: Automatic Exploit Generation. NDSS Symposium 2011, Conference paper. Retrivered from http://security.ece.cmu.edu/aeg/aeg-current.pdf, 27 Jan 2020.

[86] Shinder, D. L., Ed Tittel, Scene of the Cybercrime. Computer Forensics Handbook. (1st ed). Chapter 6. Understanding Network Intrusions and Attacks,. Syngress Publishing. 2002, p 281.

[87] *Ibid.*

[88] De Villiers (2005), *supra nota* 33, 27.

[89] Shinder (2002), *supra nota* 86, 280.

[90] *Ibid.*

[91] Rodriguez (2016), *supra nota* 28, 668.

attack, if they have access to the tools available for that purpose. This also means that the attacks can happen at a faster rate because automation can save a lot of time and effort.[92] AI that is designed to look for vulnerabilities can systematically scan available data and choose the appropriate entry point and necessary tools without input.

Another process where AI is used is attack launch.[93] The penetration of the network and the actual attack sequence can both be done automatically. As a matter of fact, the attack tools used by the attacker must not be divided into different parts. The attacker does not need port scanners, intrusion tools and attacking tools separately but the attacker can use a single tool that performs all those functions.[94] So the modern automated processes shift the control of the course of the attack from the attacker to the smith who made the tool. Those tools could be based on AI technology, able to adjust penetration and attack techniques to achieve the most effective result or they could be made by AI used by the smith.

Disguising the attacker and the unfolding attack is another important technical obstacle where automation can save the day for the inexperienced attacker. Random selection techniques to mess up patterns normally picked up by detection software, package disguise techniques to hide the nature of the traffic.[95] They can all be automated under the single attack tool that is available for the determined.

Arivudainambi and others describe cyber-attacks using AI malware. They mention the main entry ways such as vulnerabilities in networks and apps, but also transferring data through USB and through email.[96] They state that AI-based malware searches for additional entry points and "implant multiple variants in order to ensure that the attack can continue if a single point of source has been detected"[97]. This means that even if the target system detects the attack and moves to stop it, there are modified versions available that can take over and continue with the attack. They even mention that AI-created malware can delete evidence of their attack but return later to continue.[98]

[92] Shinder (2002), *supra nota* 86, 286.
[93] Arivudainambi (2019), *supra nota* 31, 50.
[94] Shinder (2002), *supra nota* 86, 286.
[95] *Ibid*., 287.
[96] Arivudainambi (2019), *supra nota* 31, 54.
[97] *Ibid*.
[98] *Ibid*.

Arivudainambi mentions several AI-specific behaviors like detecting the environment where the malware is executed, making its activity hidden in sandboxing.[99] They also mention hidden communication and evasion behaviors seen in the AI- created malware.[100]

[99] *Ibid*., 54-55.

[100] *Ibid*., 55.

2. Liability for AI-created malware

To study the impact and responsibility for the damage caused by AI, in particular, developing technologies in the field of bio-robotics on national and European legal systems, the European Commission initiated the RoboLaw project in 2012, as traditional jurisprudence cannot provide a regulatory framework, which will ensure equal rights to people in case of investigation.[101] However, this project includes only the regulation of robotic technologies,[102] but not AI-created malware, which also needs to be regulated.

However, speaking of globalization, problems related to AI cannot be determined by territoriality or be highlighting of different legal traditions practices, because they are transnational. In this regard, legal regulation cannot be applied only in Europe, since this problem also affects other countries outside the EU, which means that the AI problem should be solved at the global level with the participation of all countries of the world. The lack of legal regulation in the field of AI is a problem for the global citizenship of the entire network society, including countries where civil law and common law apply.[103]

To understand how a real investigation can look like in real life, let's look at it using the example of Estonian Penal Code, in accordance with Article 216(1)(1) “supply, production, possession, distribution or making otherwise available of a device or computer program which is created or adjusted in particular for the commission of the criminal offences provided for in §§ 206, 207, 213 or 217 of this Code, or of the means of protection which allow to get access to a computer system with the intention of committing himself or herself or enabling a third person to commit the crimes provided for in §§ 206, 207, 213 or 217 of this Code is punishable by a pecuniary punishment or up to two years' imprisonment“[104], which means any action what is related to writing or creating malware, even if it is created for the purpose of protection is a computer-related crime.

Also, Article 217 of Estonian Penal Code states: "Illegal obtaining of access to computer systems by elimination or avoidance of means of protection is punishable by a pecuniary punishment or

[101] Čerka (2015), *supra nota* 27, 377.
[102] *Ibid., 384.*
[103] *Ibid., 377.*
[104] KarS RT I, 28.02.2020, 5, §216(1)(1)

up to three years' imprisonment"[105] However, AI is different from ordinary computer algorithms, because AI can learn and generate codes by itself and can function similarly to the networks of neurons in the brain.[106] Therefore, there no existing regulation at domestic level which would prosecute AI-created malware and perpetrators who instruct AI to generate malicious codes.

To receive compensation, it is necessary to establish and prove damage, which is the main condition of civil liability. In this regard, it is necessary to establish liability for damage in the context of the legal relationship between AI and its developers. In countries with civil law traditions, the damage must be compensated by the offender or by the person who is responsible for the actions of the offender. Based on this, it is concluded that the developer is artificially intelligent and is the responsible person who will be obliged to compensate for the damage.[107]

In the U.S, the current state of the law does not criminalize or penalize for writing malware, but only for using it.[108] In this connection, it is not clear whether a malicious program can be written legally, although it will be used for legitimate and good purposes.[109] However, it should be noted that the programmer's intention is immediately apparent from the complexity of the malware.[110] In this matter, Krochinsky is rather categorical in his beliefs, but one should not forget that malicious programs can also serve positive purposes, for example, help employers to monitor actions on working computers, can provide remote assistance in case of malfunctions, etc.[111] However, Krochinski admits that the programmer-creator may not be sure of his actions and accidentally write malware, but believes that this does not change his intentions.[112]

However, when you consider that AI should get the ultimate goal. This means that if AI is hit by an end goal, it will analyze the data structure and possibly create various subtasks. Today, a computer cannot independently choose its target. This means that the programmer himself enters and defines the goal.[113] Although, the liability of the programmer will be discussed in detail later. Moreover, in should be observed that there is a possibility that in the future when the level of self-

[105] KarS RT I, 28.02.2020, 5, §217(1)
[106] Scharre (2019), *supra nota* 19, 136.
[107] Čerka *(2015), supra nota 27, 383.*
[108] Rodriguez (2016), *supra nota* 28, 671.
[109] *Ibid*., 673-674.
[110] Kroczynski (2008), *supra nota* 44, 831.
[111] Clarke, R., Maurushat, A. Passing the Buck: Who Will Bear the Financial Transaction Losses from Consumer Device Insecurity? J.L. Inf. & Sci., 18(1), 2007, p 34. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/jlinfos18&i=8 , 25 March 2020
[112] Kroczynski (2008), *supra nota* 44, 833.
[113] Davis (2018), *supra nota* 17, 185.

control of a computer rises, it will be able to determine the degree of responsibility of the people involved.[114]

Various mistakes can occur with junior specialists who just try different things, learn, and develop. There are fewer and fewer such cases, because the majority of malicious programs have already been identified. However, such errors cannot be ruled out especially when writing very complex programs, including by experienced specialists who have many critical tasks, and errors in the code for any of them can lead to an irreversible error.[115]

AI uses the principle of machine learning, which means that it can change itself and its own algorithms based on the feedback that it collects during training and use. AI has several different abilities and one of them is to create a lot of malware and to create malware that develops by itself. The United Nations Convention on the Use of Electronic Communications in International Contracts, Article 12 determines that an individual or legal entity on whose behalf the computer was programmed should ultimately be responsible for any malware created by AI.[116]

Information technologies and in particular AI are constantly changing and developing - they are very spontaneous and constantly evolve.[117] However, another problem arises here, because institutions that subsequently control the actions of AI which create a new legal framework for regulating the functioning of robot technologies should be formal, consistent and stable, because in common law systems should be legal stability and predictability supported by judicial precedents.

In this connection, should the legislation governing this area be adapted to constantly changing information technologies? Following that, the law must constantly change in order to be effective, regardless of changes in information technology, and also to be universal in order to be effective. That's why here is a legal gap. After all, the law should be relatively stable. If law changes frequently, it is impossible to understand what is right at any given moment. And in order to create a solid legal basis for the development of robotic technologies in Europe, it is necessary to give preference to fundamental legal norms and general principles of law.[118]

[114] Giuffrida, I. Liability for AI Decision-Making: Some Legal and Ethical Considerations. Fordham L. Rev., 88, 2019, p 445.

[115] Clarke (2007), *supra nota* 111, 20.

[116] Pagallo, U. The laws of robots: crimes, contracts, and torts. Springer 2013, p98.

[117] Čerka (2015), *supra nota* 27, 384.

[118] *Ibid.*

Information systems are important elements of political, social and economic interaction between member-states. In this connection, the Directive 2013/40 / EU was issued, the purpose of which was to establish minimum rules in the Union and help Member States conduct criminal law equally in the area of attacks on information systems, instead of harmonizing criminal law in the EU legislation. This directive applies only in minor cases where the damage is caused by a crime or risk to public or private interests, such as the integrity of a computer system or computer data, when the imposition of criminal liability is not necessary. Thus, Member States may apply this Directive in accordance with their national law and practice. [119]

2.1 Programmer liability

Let's start with who the programmer is and what is required for the implementation of the criminal law. Therefore, the programme**r** is a person who develops programs for different algorithms, AI, and cognitive systems[120]. Only after correct determination, we can continue analyzing the criminal law and, in particular, the liability of the programmers.

Criminal law usually requires *actus reus* which means - an action, and *mens rea* - a mental intent. Where the *actus reus* consists of an action and inaction. And the *mens rea* requires being aware, or only negligence and strict liability offences, for which no *mens rea* needs to be detected.[121]

Gabriel Halley in his book "Crimes Involving Artificial Intelligence Systems" offers three legal models by which crimes committed by AI systems can be considered[122]:

1. Perpetrator via another. If the crime was committed by a mentally disabled person, child, or animal, then the offender himself will be considered innocent because he does not have the mental ability to form *mens rea*. However, if such an agent was used and instructed by another person, then this instructor will be prosecuted. According to this investigative causation, the AI program can be considered an innocent agent, and the programmer will be considered a criminal, because he gave a command, an indication of action.[123]

[119] OJ L 218, 14.8.2013

[120] Job Wizards. Do you know what algorithm trainers do and how cognitive systems work? https://job-wizards.com/en/cognitive-systems-what-do-algorithm-trainers-do/

[121] Kingston (2018), *supra nota* 12, 270.

[122] *Ibid*., 271

[123] *Ibid*.

2. Natural- probable consequence. If the crime occurs, an AI program that was created for good purposes but was activated improperly and committed a crime. For example, in Japan, at the motorcycle factory, a worker was killed by the AI robot that worked next to him. The robot recognized the employee as a threat and decided that the most effective solution would be to simply destroy him. Using a hydraulic lever, the robot crashed the worker into another machine and instantly killed him, after which he continued his duties. Programmers can be held legally liable if they know that their onset is a natural order of events and a likely consequence of their programs or use of the application. However, this principle should divide AI programmers into those who know that the attack is being implemented or have programmed the plan to fulfil the attack goal and those who did not know and did not program the AI program for other purposes.[124]
3. Direct liability. When *actus reus* and *mens rea* are assigned to the AI system. *Actus reus* is not difficult to apply to the AI system. If, as a result of certain actions, a crime occurs or the system does not take any actions to prevent the crime, then the *actus rea* can be applied. However, it is much more challenging to apply *mens rea* because all the criteria must be met.[125]

For example, in the future, the user decides to buy a robot defender at home instead of a dog, and his software is based on AI technology. The user did not program the software, but he used the artificial intelligence program in his own interests. Such a robot will act on orders to attack anyone who enters the house. And when the robot makes an attack, the user will be considered an attacker. And such an attack will be carried out using an AI robot. When a programmer or user uses AI as a tool will be liable for preparator via another.[126] In this case, the user has a certain intention when he commands the robot to attack.[127] Accordingly, there is no legal difference between AI technology and, for example, an animal or a screwdriver is are used to achieve an unlawful purpose.[128] When a programmer creates AI malicious code to achieve certain goals, it is considered as a crime via another.

However, what to do if the programmer did not commit this and the AI system combines both the virtual and the mental elements of additional crimes, then the AI system is responsible for the first

[124] *Ibid*.
[125] *Ibid*., 272.
[126] Hallevy, G. Liability for Crimes Involving Artificial Intelligence Systems, Switzerland: Springer, 2015, p110
[127] *Ibid*., 111.
[128] *Ibid*., 112.

type of responsibility. However, anyway, the question arises of the criminal liability of the programmer. The third is responsibility for the possible consequences that were committed in practice, but these crimes were not part of the original unlawful plan.[129] In general, the issue of liability for possible effects refers to the criminal liability of one person for planned crimes committed by another person.[130] And perhaps in this proposed model the most appropriate criminal liability. But what if it turns out to be an unplanned crime?

For strict criminal liability is not necessary to establish the factual and psychological elements unplanned crimes and participate in any subsequent criminal act subject to criminal prosecution. The main idea of this approach is to prevent the participation of potential criminals in future crimes by expanding criminal liability not only for planned crimes, but also for unplanned ones.[131]

A potential attacker must understand that his or her personal criminal responsibility cannot be limited only to specific types of crimes and that he or she can be prosecuted for all expected and unexpected events that are directly or indirectly related to his or her behavior. Criminal liability for an unplanned crime is imposed regardless of the role of the offender in the commission of the planned offense as a criminal, instigator or accomplice. Therefore, the criminal liability for an unplanned crime is the same for all parties and does not require real and mental skills. However, the western legal systems consider such a restraining approach too extraordinary and therefore do not apply it.[132]

The criminal liability for the use of AI is divided into two types. The first type will be used when the programmer decides to commit a certain violation, but the system surpasses the plan and commits other crimes or does additional harm. The second type concerns cases when the programmer did not program the artificial intelligence system to commit any violations, but the system itself committed a crime.[133]

In the first case, criminal liability is divided into liability for planned violations and unplanned violations. If the programmer planned that the system would commit a specific violation in particularly a written malware, the most important thing in this case the programmer is fully liable

[129] *Ibid.*, 113.
[130] *Ibid.*
[131] *Ibid.*, 139-140
[132] *Ibid.*, 115.
[133] *Ibid.*, 118.

for this violation, as in the case with a screwdriver or a dog.[134] Moreover, if additional unplanned damage was a likely consequence of a planned violation, the programmer will be prosecuted for the unplanned violation in addition to criminal liability for the planned violation.[135]

The second type will be considered an accident. In this case, the programmer will be liable only for possible consequences, and such responsibility is intended to combat unplanned consequences and unplanned crimes. However, using such a mechanism to eliminate faults in accidents without criminal intent would be disproportionate and too harsh.[136]

However, the criminal liability of the programmer for an unplanned crime will be considered separately, since the programmer was not going to commit any violation, and the mental element of this intention is not important. The criminal liability of the programmer in this case should be considered as negligence.[137]

2.1.1. Creation of AI for malware manufacturing

Immediately after the 1988 attacks on U.S research centers, military concern about computer viruses became widespread, leading to the proposing of The Computer Virus Eradication Act of 1989 by the Congress, this included turning computer virus writing to a criminal offense.[138] The previous Acts that came out in 1984 and 1986 already included provisions for punishing some form of computer crimes.[139] During those times, trojan horses, logic bombs and computer worms were the main types of malware.[140] In terms of programmer liability, those early times can offer us clues for the liability related to malware. Concerning the 1986 legislation, Hansen points to the wrong mental state of the criminal *mens rea* applied.[141] Hansen argues there that when the virus is planted, the programmer loses control over target selection and attack effects and that intention or knowledge cannot be applied.[142] However, the Computer Fraud and Abuse Act (CFAA) does forbid negligent damages caused by illegally accessing a computer that is protected.[143] It also

[134] *Ibid.*
[135] *Ibid.*
[136] *Ibid.*, 119.
[137] *Ibid.*
[138] Hansen (1989-1990), *supra nota* 29, 718-719.
[139] *Ibid.*, 730-731.
[140] *Ibid.*, 721.
[141] *Ibid.*, 734.
[142] *Ibid.*, 734.
[143]18 U.S.C., §1030(a)(5)(B). Retrieved from: https://www.govinfo.gov/content/pkg/USCODE-2010-title18/pdf/USCODE-2010-title18-partI-chap47-sec1030.pdf, 7 May 2020

seems to punish any damages caused in such way, even without negligence or intention as long as the access to the computer was intentional.[144] Oddly, only the computer used by a financial institution, or the U.S government, or a computer used in commerce or communication at interstate or international level is considered a protected computer by the Act.[145]

According to Rodriguez, the current state of the law in the U.S does not make malware writing punishable, just its use.[146] Rodrigues then shows how the current provisions in CFAA have been used against cybersecurity researchers and that this is basically the same as punishing writing.[147] She offers an interesting comparison where she compares the position of a malware writer with the position of a gun manufacturer.[148] Eventually, Rodriguez finds that it is still unclear if malware can be legally written, although malware is often used for legitimate purposes.[149] Kroczynski writes that the intent of the programmer can be seen from the complexity of the malware.[150] In this sense, Kroczynski is quite firm and considers intent to be clear. But malware can also serve positive purposes like helping employers monitor activities on work computers, allowing remote assistance in case of malfunctions etc.[151] However, Kroczynski also thinks it is possible that the author of the virus releases it accidentally through network or with connected programs.[152] He also sees the possibility that the author may be unsure about the functionality of the written malware but thinks that this does not change the intent.[153]

Davis brings an interesting point to the discussion. He claims that AI must be given the end goal. That means that there could be sub-goals that the computer creates because of the data pattern analysis, but the end-goal is given to AI.[154] The computer is not autonomous to select its own goal. This means that the programmer must write the goal into the original algorithm. It may change in the future, but right now, the programmer who writes this end goal into the original program, is the one who would be responsible for the damage that the program does. In the future, it can

[144] 18 U.S.C. § 1030(a)(5)(C). Retrieved from: https://www.govinfo.gov/content/pkg/USCODE-2010-title18/pdf/USCODE-2010-title18-partI-chap47-sec1030.pdf, 7 May 2020

[145] 18 U.S.C. § 1030(e)(2). Retrieved from: https://www.govinfo.gov/content/pkg/USCODE-2010-title18/pdf/USCODE-2010-title18-partI-chap47-sec1030.pdf, 7 May 2020

[146] Rodriguez (2016), *supra nota* 28, 671.

[147] *Ibid*., 671-673.

[148] *Ibid*., 673.

[149] *Ibid*., 673-674.

[150] Kroczynski (2008), *supra nota* 44, 831.

[151] Clarke (2007), *supra nota* 111, 34.

[152] Kroczynski (2008), *supra nota* 44, 833.

[153] *Ibid*.

[154] Davis (2018), *supra nota* 17, 185.

happen that the level of self-control that the computer has can determine the extent of liability of the involved humans.[155]

2.1.2. Creation of faulty AI

One problem regarding programming is the accidental creation of something not intended. If a programmer is skilled and working according to a predetermined plan and a goal to achieve, it is not likely that something completely different will be created. But when a junior programmer is playing around with different tools, they may end up creating something bad. In early malware related legislation, Hansen found a problem with proving the programmer's knowledge that the program is a virus and understanding that this program is harmful.[156] He also brings in a possible amendment to the *mens rea* requirement, changing intention threshold to negligence.[157]

In the early days of computer programming, it was possible that the programmer who experimented with new commands and functions would accidentally create a virus. In the present day, this can happen too, but mainly with the junior programmers who just try different things. It can also happen when very complex programs are written that have many critical tasks and errors in the code for any of those can result in a critical error and damage to the user. Although it is rare for experienced programmers to produce known types of malware without realizing it, other errors in software writing do happen, bugs do exist. Some have even written that security patches meant to cover vulnerabilities often contain bugs creating new vulnerabilities too.[158]

However, if a faulty program causes damages, the author must examine liability for those damages. One way to do this is through product liability. If a client has bought a commercial software or has commissioned some software from the programmer, the client may wish to use product liability to get compensation. The first question that the court could ask is whether software is a product or a service.[159] According to Ningsih, the main problem for applying product liability to software is the intangible nature of computer programs, although Ningsih then

[155] Giuffrida (2019), *supra nota* 114, 445.

[156] Hansen (1989-1990), *supra nota* 29, 739.

[157] *Ibid*., 740.

[158] Clarke (2007), *supra nota* 111, 20.

[159] Ningsih, A. S., The Doctrine of Product Liability and Negligence Cannot Be Applied to Malware-Embedded Software. *J Indonesian Leg. Studies*, Vol. 4, No. 1. 2019, p 9. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/jils4&i=11, 8 April 2020

concludes that product liability has been widened to intangibles like gas and writings.[160] Product liability includes liability for damages that are caused by the defective product.[161] Now, we need to examine if software could be under product liability.

Generally, there are three types of product liability available. Those are strict liability, negligence, and the third one is a breach of product warranties.[162] Strict liability occurs anytime when a product causes damage, the producer's intentions or negligence does not matter.[163] Negligence in the software world could occur through the defective testing by the programmer.[164] To establish negligence, some criteria must be satisfied. The producer of the software must have a duty of care towards the injured person.[165] The producer must breach that care, and this must cause damages for the negligent liability to arise.[166] According to the Liability Directive 85/374/EEC, the principle of liability without fault applicable to European producers. Where a defective product causes damage to a consumer, the producer may be liable even without negligence or fault on their part.[167]

Ningsih considers the negligence liability and warranty-based liability to be a better option compared to strict liability, because strict liability can expose programmers to penalties even if they are cautious.[168] And most likely this is the most reasonable option to date. Second problem with liability is that in many cases, software producers do not offer any warranties and severely limit their liability through end-user license agreements (EULAs),[169] something that the state of the trade practice regimes make worse.[170]

[160] *Ibid.*, 10.
[161] *Ibid.*
[162] *Ibid.*
[163] *Ibid.*
[164] *Ibid.*, 14.
[165] *Ibid.*
[166] *Ibid.*
[167] OJ L 210, 07.08.1985
[168] Ningsih (2019), *supra nota* 159, 18.
[169] Kirtley, J. E., Memmel, S. Rewriting the Book of the Machine: Regulatory and Liability Issues for the Internet of Things. *Minn. J.L. Sci. & Tech.*, Vol. 19, No. 2, 2018. pp 507-508. Retrieved from: https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1448&context=mjlst 7 May 2020
[170] Clarke (2007), *supra nota* 111, 24-25.

In identifying defects and determining the damage caused by these defects, as a rule, the programmer's negligence or the developer's negligence is established, but criminal liability does not occur.[171]

In order to file a malpractice lawsuit, the American lawyer Maruerite Gerstner[172] identified three necessary elements[173]:

1) a duty of care;
2) a preachment of that duty;
3) a caused damage.

Gerstner has analyzed the negligence of strict liability on the example of the U.S law. This responsibility is applied when the products are defective and dangerous during use and ultimately cause physical injury. If such law is applied to developers of algorithms and the AI systems, they are obligated to ensure that their systems do not have defects that can be dangerous to human life.[174]

With AI, things can get more complicated. AI uses the principle of machine learning. It is not a static program written by a programmer, but it can modify itself, it can change its own algorithms based on the feedback it collects through training and use while ordinary computer programming can produce errors that sometimes can have very bad consequences. But the errors in AI writing can be worse. The ability of AI to modify itself makes errors more likely. If a mistake is made during the programming of AI, a good purpose can lead to a bad result. For example, if an AI system is created to produce software that gains access to certain networks or systems to modify or update data, there can be a legitimate purpose. But mistakes in building AI can lead the system to change the software it produces.

Kroczynski has analyzed *U.S v Morris* case, from the current legal point of view.[175] The case concerns a Cornell University student who created a worm with a purpose to expose security flaws but which caused unintentional damages.[176] Kroczynski finds that although under the new version

[171] Tuthill, G.S. Legal Liabilities and Expert Systems, AI Expert. 1991. In Kingston, J. Artificial Intelligence and Legal Liability. Australian Product Liability Reporter, 28 (3), 2018. p269-279. Retrieved from https://arxiv.org/pdf/1802.07782.pdf , 25 March 2020

[172] Gerstner M.E. Comment, Liability Issues with Artificial Intelligence Software, 33(1), Santa Clara Law Review, 1993. p239. Retrieved from: http://digitalcommons.law.scu.edu/lawreview/vol33/iss1/7, 25 March 2020

[173] Kingston (2018) *supra nota* 12, 273.

[174] *Ibid*., 273-274.

[175] Kroczynski (2008), *supra nota* 44, 838.

[176] United States Court of Appeals, Second Circuit, 928 F.2d 504, *U.S. v. Morris,* 07.03.1991

of the law, intention to cause damage must be shown, the student would still be tried for negligent damage and unauthorized access by 18 U.S.C. § 1030.[177] A big part of the liability would rely on unauthorized use of the protected computers (governmental).[178] But what about the computers of people and businesses? As of 2008, Kroczynski found the New York law to be similar to federal level, same with New Jersey, but the Pennsylvanian law outlaws possession of malware if there is desire to spread the malware.[179] He proposes to make writing and also the possession of malware into a criminal offence, that would apply to the average user.[180] Still, he understands that there might be legitimate uses of malware for antivirus programs.[181] His approach is interesting, especially the term average user. Who is an average user? Is it a child doing homework, a teenager interested in programming, an employee in a retail sector, an employee in a software company, or a computer programmer, AI researcher? This is the first issue with Kroczynski's approach. It is difficult to define the average user because we are all so different, we have different interests, motivations, skills. Defining the average computer user is impossible. The second problem is related to the definition of possession of the virus and the writing of the virus. Code in a high-level programming language can be written in a text form, and compiler software is easily accessible. If the user's computer has been infected with malware, would it always be possible to tell if the person wrote the code or it had spread from somewhere else? The current hiding capabilities of viruses makes it more complicated.

2.2 Trainer liability

Algorithms are a sequence of computational steps that are used to solve computational problems by writing computer programs, and when we describe algorithms using a computer programming language, it is called a program.[182] Therefore, in other words, an algorithm is a formula or calculation for a computer whereas a trainer is a person who provides a self-learning system with useful feedback.[183]

[177] Kroczynski (2008), *supra nota* 44, 838.

[178] *Ibid*., 838.

[179] *Ibid*., 840-842.

[180] *Ibid*., 848.

[181] *Ibid*., 855.

[182] Joshi K. B. Data Structures and Algorithms in C++, Tata McGraw Hill Education Private Limited: New Dehli, 2010, p236

[183] Job Wizards. Do you know what algorithm trainers do and how cognitive systems work? Retrieved from: https://job-wizards.com/en/cognitive-systems-what-do-algorithm-trainers-do/ , 7 May 2020

Programmers write programs for AI; however, the development of certain algorithms requires the work of a trainer.[184] The mission of the AI trainer is to expand AI beyond its technical capabilities by adding to it the human element of reason, personality, and empathy. That is, as mentioned above, to make AI a powerful machine from a simple machine.

AI or machine learning algorithms are based on two learning methods: supervised and unattended. The first requires a coach. The second one works without a trainer; the system independently searches and determines patterns and filters them. For example, such a system is used in Amazon when a person recommends shopping, and such recommendations are based on previous experience and the interests of other buyers.[185]

Input data is vital for such a system training, because if there is a system malfunction or the data is outdated, it can critically affect the efficiency and quality of work of AI.[186] However, despite the autonomy and independence of the system, in any case, it should be controlled by trainer.

In order for AI to do or study only what the trainer wants, it is necessary to enter data from a set of training data. For example, the trainer can enter data for autonomous vehicles that will allow you to assess damage and take actions based on the well-being of the owner of the vehicle.[187]

If the computer is able to determine the parameters that need to be taken into account in order to make a specific decision, in this case, the goal of the trainer can be easily foreseen. Of course, we can say that the trainer could not predict certain consequences and that it was an accident, but this is only possible if the input data are common, in which case the responsibility of the trainer will be difficult to prove.[188]

If the trainer enters only malware as input, then the intention will be apparent, and accordingly, the responsibility will be determined. However, for cybersecurity or forensic purposes, the trainer may declare that he is developing new tools to combat malware. Also, in defense of the trainer, it can be said that AI has not yet been well studied and it is difficult to predict its behavior because

[184] *Ibid.*
[185] *Ibid.*
[186] *Ibid.*
[187] Davis (2018), *supra nota* 17, 180-181.
[188] *Ibid.*, 181-182.

it changes its own algorithms.[189] And in this case, the trainer's responsibility will definitely decrease. But if the trainer accidentally includes various software in the source code and skips the malicious program in the data set who does this makes his action sloppy.[190]

2.2.1. Purposeful mistraining

Modern AI technology relies on neural network and deep learning through data analysis. Incoming past data and trainer or programmer instructions are the input and a decision is an output. But the pattern recognition and algorithm making does not just happen, and it needs training by the past data. If AI is mistrained purposefully, then there can be very bad consequences. Because we are dealing with malware made by AI, we must ask what is the liability for intentionally training AI to create malware?

We should start from other AI applications. Making a working autonomous vehicle requires a lot of training and a huge amount of data input. But it can also require making decisions about data not received from sensors. Davis mentions the decision-making by autonomous vehicles when they are required to limit accident liability.[191] This parameter can be directly programmed into the original algorithms or they can be achieved by training. Because manufacturers would think of legal liability, they want AI in the vehicle to reduce that liability as much as possible. Therefore, they would use datasets in training that in case of accidents, AI would use cost-benefit analysis. They would encourage AI to choose the least costly solution. But as Davis has mentioned, this means that the vehicles will put price tags on people and property and the rich will be safer.[192] In this case, AI is purposefully mistrained to prefer saving the rich and powerful and destroy least property if an accident happens. Data can be manipulated. When the trainer does not want AI to learn something, he or she can leave that data out from the training dataset. Then AI will only learn what the trainer wants it to learn. In autonomous vehicles, this means possibly wealth-based discrimination in accident cases. In the case of software-making AI, this means possibly malware writing by the computer.

[189] *Ibid*., 182.
[190] *Ibid*., 187.
[191] *Ibid*., 180.
[192] *Ibid*., 180-181.

There is an important question asked by Davis about AI liability. He asked if the computer is able to tell what parameters it took into its consideration when making a decision.[193] If the computer is able to do that, the intent of the trainer may be more easily seen. If not, then it is possible for the trainer to hide it. When the trainer of autonomous vehicles must answer questions about the training data given to the AI, then the trainer can say that he or she did not think about that outcome, that it was an accident. This is possible if the discriminatory data is common. If that happens, it is difficult to prove the trainer's intent and liability.

But what about malware creation? If the trainer gives the computer malware as input and the computer produces new malware, then the liability must be clear? It may not be so. If the trainer gives the computer only malware as input, then it is likely that intent is shown, and liability can be proved. But if the trainer gives the computer some software as input which might include malware or if the input software samples have some properties related to malware like cookies, adware, spyware that are also used legally, then the result can be different. The government also uses malware and similar tools for national interest purposes and digital forensics can use malware too. Cybersecurity companies can produce new malware to develop new defense tools. It is possible for the trainer to claim that they wish to develop new tools for cybersecurity or forensics industry and that they used malware as input for that purpose.

Another defense for the trainer is that the results of AI operation are not completely clear still. There is a lot that cannot be predicted well about how the AI solves problems because it changes its own algorithms.[194] This means that the trainer might get away with mistraining, or at least the trainer's liability is reduced. When we come back to Davis's argument that AI needs to be given the end-goal, the conclusion is valid here too. If the end goal is not given in detail by the programmer, the trainer may be able to set the goal. When the trainer sets malware creation as an end-goal then the computer will do so. The trainer can give AI some hints or parameters that the computer uses to create a program that turns out to be malware. Davis understands the possibility of changing the moral decision-making of AI through input data.[195] But he understands that there are many things that influence decision-making in the input data.[196] Here we can see that it is possible for AI to begin producing malware even if it was not programmed to do so because of

[193] *Ibid*., 181-182.
[194] *Ibid*., 182.
[195] *Ibid*., 187.
[196] *Ibid*.

the training it received. But the results of the training are not always perfectly clear. If malware is the only thing used in training, then the trainer would have liability. But if there is various software included in the training dataset (including malware) and the computer creates different programs, some of which turn out to be malware then the liability is not that clear. The trainer might have missed the malware in the dataset, making his or her actions negligent (which we will discuss in the next part). Or the trainer wanted to make a program for good purposes that uses some qualities of malware like remotely changing computer settings, for example. Another intentional malware creation would result from hacking the system from outside. Also, if AI is given bad data from outside, this can ruin the results too.[197] Scharre even writes that sometimes, the attacker input is not related to the training data, but no good protection measures exist to prevent such attacks.[198]

2.2.2. Negligent mistraining

Liability for use of malware in training process due to negligence is a potentially important problem that we must discuss. During training, many things can go wrong and the AI may pick up patterns it is not supposed to. It is possible that poor training practices can make AI think that it must create malware. Because this mistraining can have bad consequences, we must see how the liability would work in mistraining situations.

Training is necessary in all AI applications. The machine needs to identify all available parameters and the end result and redesign its own algorithms to achieve a desired result. This happens through training. Data is fed to the AI and the AI must find patterns in this training data to produce a similar result in similar circumstances. It must be able to develop rules and principles to be written into its algorithm based on the patterns found in training data. AI can create new variables based on input data even without the trainer's knowledge what these variables are.[199] Therefore, we can see that the training phase is very important in AI development. In those cases where AI is written well but the training data is incorrect, the resulting AI may perform badly or even be harmful.

[197] Scharre (2019), *supra nota* 19, 141.
[198] *Ibid*., 141-142.
[199] O'Donnell, R. M., Challenging Racist Predictive Policing Algorithms Under the Equal Protection Clause. N.Y.U. L. Rev., 94, 2019, p 551.

Training is especially important in autonomous vehicles. Currently, the technology of autonomous vehicles is dependent on mapping data and identifying certain signals related to road signs, other vehicles, pedestrians, road limits. The vehicle collects data from the surrounding environment through many sensors and compares it with a previous map, making the map more accurate.[200] For this to happen, the vehicle must first know how to behave according to an existing map. This is done via training. The trainer can first feed data to the vehicle AI in a simulation environment to create first patterns for recognition. Later, the trainer can take the vehicle for a drive where the signals collected from the environment are compared to the previous datasets, and driver's behavior is compared to the previous behavior of AI. This is the basis of autonomous vehicle AI training.

But what would happen if the training does not go very well? There can be many reasons for this. Those who drive cars do not all drive the same way. Sometimes, the AI that is trained by several trainers can receive confusing datasets. That can lead to unwanted behaviors. It can also happen that the trainer gets distracted, signs are covered by leaves or snow, visibility is low, or something unexpected happens during the training session. In that kind of situation, some principles may be broken, and wrong patterns may be reinforced.

Training issues related to deep learning do not exist only in autonomous vehicle technology, but they can happen in other areas too. As I wrote before, deep learning means that the computer is given datasets with conditions and outcomes and the computer then uses its deep neural network to find patterns and create algorithms that are able to simulate those patterns. But training does not always lead to good results. One AI that was trained to recognize clothing performed very bad.[201] Those who train AI can do so with two ways. They can train it with made up data or real-life data. Made up data is more difficult to use because the amount of data that can be simulated is very small and it can be inaccurate. Real life data is most used. It takes previous situations and the circumstances and also outcomes and it forms decision-making algorithms based on this data experience. But there can be dangers with this sort of approach. Davis has found that the past data can give AI bad influence and this can contribute to the deepening of discrimination in employment, for example.[202] Medical school admissions and reoffending rate prediction have also

[200] Kemp, R., Autonomous Vehicles – Who Will Be Liable for Accidents? *Digital Evidence & Elec. Signature L. Rev.*, 15, 2018, p 40. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/digiteeslr15&i=37, 15 Feb 2020.

[201] Bedeli, M., *et al.*, Clothing Identification via Deep Learning: Forensic Applications. *Forensic Sci. Res.*, 3(3), 2018, pp 219-229. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/forsr3&i=221, 25 March 2020

[202] Davis (2018), *supra nota 17*, 179.

shown bias.[203] He brings an example where he says that a computer may consider women to perform more poorly than men in similar circumstances, discriminating them because there was a pattern of discrimination before.[204] Thus, this means that the computer can make discriminating decisions based on past data because discrimination was very widespread in the past. This is one example of how bad training data can influence the outcome of deep learning. The computer does not have the emotional intelligence or understanding of morals to decide which past experience is good and which is bad. It only works on data analysis.

Davis asks important question related to employment discrimination by AI that is also important for AI based malware liability. He asks if there is any discriminatory intent present for data mining and AI analysis. We have already analyzed the intentional mistraining of AI. But unintentional mistraining could still lead to negligence. Some patents related to bias reduction in AI have been applied for and granted too, but there have not been specific descriptions how AI is trained to reduce that bias.[205] Cofone has found that there can be self-selection of biased data in the sense that the biased data may cause the change in algorithm that then confirms this change with new biased data.[206]

2.2.3. Training faulty AI

Possibility of liability related to training a faulty AI system is also a possibility that we must discuss. As we already mentioned, programming errors happen all the time and there are no guarantees that the development process of AI does not have any errors. Therefore, the question is who is liable for training AI with flaws in it? Is it the programmer who made the error, or is it the trainer?

If the programmer knows that there is an error, then the programmer would definitely have product liability issues, as we have mentioned earlier. But the programmer might not have knowledge about the problems. What would happen if the trainer discovered the flaws but decided to use AI anyway? There could be reasons for this behavior. AI development takes a long time, and the training can take a long time too. If the trainer would send AI back to the programmer, it can delay

[203] Sloan, R. H., Warner, R., Algorithms and Human Freedom. *Santa Clara High Tech. L.J.*, 35, 2018, pp 22-24.
[204] Davis (2018), *supra nota* 17, 179-180.
[205] Violago, V., Quevada, N. AI: The Issue of Bias. *Managing Intell. Prop.*, 227, 2018, p 34. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/manintpr277&i=34, 25 March 2020
[206] Cofone, I. N. Algorithmic Discrimination is an Information Problem. *Hastings L.J.*, 70, 2019, p 1403.

operations for a long time and cause a lot of damage. It can also happen that the trainer sees the error but does not consider it to be a serious one. What liability does the trainer have in that situation?

If there is an error but the trainer thinks it is irrelevant then it is necessary to do a foreseeability analysis. If the creation of malware is not foreseeable, then the trainer should probably not have much liability because the trainer has not been negligent. But the problem is that different jurisdictions can define foreseeability different and that can mean that the outcomes are different.[207]

If the trainer has no knowledge of the error, then there is very little chance of liability for the trainer. The trainer would not be able to determine the result of the data analysis that the machine makes. Latent machine bias is mentioned as a common thing happening when the algorithm puts incorrect meaning to some data pattern it finds and uses it to make decisions[208]. The unreliability of AI outside its comfort zone has been seen before.[209] Scharre writes that AI can sometimes use wrong methods to make decisions, even if the decisions are aimed at the end goals.[210] AI often seems to cut corners to achieve its tasks in surprising ways, as Scharre showed.[211]

2.3. Operator liability

Something that companies need to keep an eye on is their internal security. This is especially so with those who operate AI. It is possible that employees who will be laid off, disciplined, passed for promotion, lured by competitors, who have personal issues with their colleagues or executives could use their access to AI to make AI attack the company itself. This purposeful attack by insiders should be taken into consideration.[212] In this case, two liability paths can be seen. It is very clear that if the attack was made purposely by an employee of the company, this employee is liable for the attack. If the employee instructed AI to create malware that is then used to target the company, this employee has full liability. This works in theory, but what about practice? The

[207] Swanson, G. Non-Autonomous Artificial Intelligence Programs and Products Liability: How New AI Products Challenge Existing Liability Models and Pose New Financial Burdens. *Seattle U. L. Rev.*, 42, 2019, pp 1216-1218.
[208] Violago (2018) *supra nota* 205, 34.
[209] Scharre (2019) *supra nota* 19, 135.
[210] *Ibid*., 140.
[211] *Ibid*., 140-141.
[212] Shinder (2002), *supra nota* 86, 282.

defense in court could argue that although the employee was responsible for the attack. But the company could be partly liable too. This can happen if the company does not have good internal security practices. For example, this can happen when the logs of people who can access AI are not collected. Or it can be that these logs can be manipulated. Connecting AI to the company's internal networks can also be bad if it can be avoided. It would be good to have some supervisory system in place and proper network security measures. If the company does everything, they can to protect their network, then the liability is on the employee. But if the company is not making effort to defend their own systems, this can reduce employee liability. That will possibly reduce the compensation that the company gets. The same can happen when the operator accidentally gains access to parts of the company's networks that were not properly guarded. The effect of data not meant for use by AI could mean that AI develops a new algorithm that can result in damage to the company or third parties. Company's internal security can also be used as a way to keep away the curious minds inside the company. Occasionally, those who want to see how things work may change different settings for their computers and network, that can accidentally give them access to information that they should not be able to gain access.[213] Therefore, we can see that a person with no bad intentions can accidentally discover AI and modify it just to see what happens. The chance that someone accidentally instructs AI to produce malware like that is small, but it can happen. If it does happen, then the employee has been negligent in actions, but the company has created the conditions that made it happen. The company has been passively negligent. A big part of liability would then be on the company exploiting AI as it left its interfaces too easy to access.

Another liable party could be the consumer. This relates to two problems. The first is the proper security measures that the consumer uses on his or her devices, and the second one is the downloading and installing malware.

The security measures that consumers take to protect their computers, smartphones, and other Internet-accessing devices are very important in the liability analysis. But what happens to the liability when the user downloads and installs malware? Here, we can divide the problem into three parts. The consumer may intentionally get malware on his or her computer, but this can also happen unintentionally when the consumer is negligent, or it can happen when the user has taken protection measures.

[213]*Ibid*.

The user may get malware on their computer intentionally. It is hard to imagine that anyone would want to attack themselves, but it is possible. The human mind is unpredictable. For example, it may happen that a person may want to launch a large attack or wants to see some malware get spread. Then they may want to hide that they started this attack by making themselves as the first victim. In that case, the liability for damage would be on this consumer who downloads and installs the malware. If that consumer starts the fast spread of the malware that would not happen without their actions, then it may be that this consumer would have the majority of liability. Then the consumer would have a bigger liability than the programmer because the consumer released the malware.

Negligence is often a big reason why malware gets downloaded and installed on people's computers. People who visit pornographic websites, open every email attachment, follow every link, visit deep web, use pirated software are very vulnerable to malware. When people do that, they may download content that they are not even aware of.[214] In that case, the inserted malware is installed unintentionally and even without the knowledge of the consumer. It all happens in the background where it is not visible. Therefore, what would happen to liability in this case? Clearly, the consumer has not been careful, because they have done high risk activities like visiting suspicious websites and opened links and attachments that they do not know if they are dangerous or not. Like before, the liability of the consumer, in this case, relies on the foreseeability of such danger. In the early days of the Internet, when the knowledge of malware was not very good among consumers, those consumers who visited the many corners of Internet might not have been aware of the dangers. The threat was not foreseeable for them. But now, these threats are known well. Knowledge of threats and failure to increase security reasonably can make the consumer liable at least partially if someone suffers damage.[215]

However, things can be different if the person thinks that they are visiting a trusted website. Some phishing scams direct the user to a website that looks like a legitimate website and has the URL that looks identical to the real website at first. It may be masked, or it may contain a small error that is hard to see. What would happen to the liability of the consumer in that case? It can also happen that the consumer wishes to download a webpage html file but, on the background, executable file is downloaded too.[216]

[214] Clarke (2007), *supra nota* 111, 26.

[215] *Ibid.*, 46.

[216] *Ibid.*, 32.

2.3.1. Purposeful task of writing malware

As the author mentioned already in the programmer liability and trainer liability part, AI needs to have the end goal to work with. This means that someone needs to give the computer some goals. If it is not the programmer or the trainer, then the operator is the one who provides the computer with its objectives. Therefore, how does the liability work if operator tasks AI to create a malware program?

AI can be seen as a severe source of danger, and a programmer or a manufacturer on whose behalf it acts should be held accountable. Thus, it would be useful to use the theory of "deep pocket" to update the General Product Safety Directive 2001/95/EC[217] and the Product Liability Directive 85/374/EEC[218] to ensure a software manufacturer's liability. Therefore, it means that a manufactory engaged in harmful activities that are profitable and beneficial to society must compensate for the caused damages to affected people from the received profit.[219]

2.3.2. Rogue AI

AI can create malware on its own, without command to do so but as a goal it defined for itself or as a tool to reach the true goal it was given. If this happens, is there some liability of oversight? As Davis has written, it can be impossible to check the thought process of AI because using a human for this takes too much time.[220] And if computers did that, then „It may not be possible to superimpose yet another layer of analysis ".[221]As Terry has written, the amount of data that passes through AI makes it difficult to see how it achieves results and trust towards AI may replace checking.[222] However, this is very dangerous. In this case, there is no way of checking the analyzing process of AI and everybody just presume that the AI is right. But if AI is considered to never make mistakes then who to blame if mistakes happen? Then one of the people working with AI could have liability. It could be the programmer, trainer or operator.

[217] OJ L 11, 15.01.2002

[218] OJ L 210, 07.08.1985

[219] Čerka (2015), s*upra nota* 27, 387.

[220] Davis (2018), *supra nota* 17, 182.

[221] *Ibid*.

[222] Terry (2018), *supra nota* 10, 160.

This causes very serious problems. To establish legal liability, it must be made sure that the person has some sort of obligation, and that obligation was broken. This analysis can only be done by looking at the evidence. But if the computer processes so much data that it makes it impossible to determine the AI's decision-making process, then there would be no evidence. The amount of data is just too overwhelming. In that case, nobody could be connected to the malfunction. Then it is impossible to see who caused AI to produce malware and how it happened. No liability could be made. Only perhaps the liability of a company if the programming, training, and operation all happen within that company. Then there could be product liability or service liability of the company as their product or service is malfunctioning.[223]

According to Article 11 of Directive 2013/40/EU in addition to the necessary measures, Member States must ensure that legal persons can be held liable for offences referred to in Articles 3 to 8 of the Directive, committed for their benefit by any person, acting either individually or as part of a body of the legal person and take effective sanctions to a legal person and insure his liability, which include criminal or non-criminal fines, such as[224]:
(a) exclusion from entitlement to public benefits or aid;
(b) temporary or permanent disqualification from the practice of commercial activities;
(c) placing under judicial supervision;
(d) judicial winding-up;
(e) temporary or permanent closure of establishments which have been used for committing the offence.

However, it can also happen that even the company cannot be liable. The real world can sometimes have bias and this bias is entered into AI even if the training was done with unbiased data.[225] Data collected from the world can be biased, inaccurate, or incomplete and that can cause problems with AI function too.[226]

[223] Giuffrida (2019), *supra nota* 114, 442.
[224] OJ L 218, 14.8.2013
[225] Giuffrida (2019), *supra nota* 114, 442.
[226] Valentine, S., Impoverished Algorithms: Misguided Governments, Flawed Technologies, and Social Control. Fordham Urb. L.J., 46, 2019 , pp 387-393.

2.3.3. Responsibility for network integrity

Network integrity plays a major role in the liability game. Leaving a network poorly guarded is like leaving the front door of the house wide open so that the thief could come in. As I already discussed before, one of the most important parts of the attacks against users or companies is finding the vulnerabilities of networks and systems and gaining access. If the door is left open, the entrance has no obstacles. Poor computer and network integrity practices can lead to very serious consequences. They can contribute to the success of the attack and increase the magnitude of damage. Poor security can contribute to the attack in several ways. First, reckless employees can use the company's computers to gain access to the external network. They can use the computers for personal reasons or search for work-related information on the Internet. Doing so, they may stumble on infected websites that download malware into their work computer.[227] They usually do not even know that it happens. The same thing can happen if they open email attachments coming from strange addresses or from seemingly familiar ones.[228] After downloading and installing, the malware can download additional tools and malware without the employee's knowledge. The malware can also gain access to the company's internal network through the infected computer and use that network to distribute itself and other malware to other systems.

If AI is involved in creating the malware, then the employees may create opportunities for creating specific malware that is very effective in their system. If we assume that after entry, AI attack tool monitors the network and the systems connected to it, AI can use machine learning to develop a specific tool to most efficiently penetrate the entire network, carry out its attacks and leave without a trace. This can cause an attack that is multiplying itself. It can even turn the target company's entire network into a zombie and use its systems as a group of agents to gain access to the business partners' networks. Even very early on, legislation against computer crimes included liability for victims of computer crime who became knowledgeable about the crime but continued to allow using their system for further crimes.[229]

A big part of network security is collecting logs on network traffic and parameters too. As said by Arivudainambi, the log trails and advanced analysis tools can help with investigation.[230] If those

[227] Shinder (2002), *supra nota* 86, 287.

[228] *Ibid.*

[229] Hansen (1989-1990), *supra nota* 29, 736.

[230] Arivudainambi (2019), *supra nota* 31, 54-55.

log trails are not in place and if network traffic cannot be analyzed, then it may be impossible to even identify the attack. It is also much more difficult to catch the attacker. The design of computer systems, network, software and operation systems used are all important for maintaining security. What happens if the person responsible (we shall call that person administrator) for the security of network and systems creates vulnerabilities or leaves them unfixed? De Villiers considers that the administrator's liability for the breach and damage would result from negligence. Still, there is a possibility that this liability is ended if there is an intentional attack that uses this vulnerability.[231] But in such cases, the economy of lawsuit starts to play important role. Usually, the hacker does not have the ability to compensate the victim and if the liability of the administrator is no longer valid, then the victim might not receive any compensation at all.[232] To deal with this issue, it is helpful to turn to Encourage Free Radical (EFR) doctrine which keeps the liability of the administrator who has encouraged the free radicals to act by being negligent in providing protection measures.[233] Under the EFR, the attackers are considered free radicals because they are protected against liability due to age, financial situation, mental capacity, or other reasons.[234] The administrator who created vulnerabilities that were used by the attacker or administrator who did not update the security of systems or network is liable under this doctrine.[235] De Villiers has discussed how the poor security standards of companies have made them liable for damages caused by those free radicals.[236] The EFR doctrine concerns civil liability. But here we also deal with criminal liability. But the situation is not much different. In many jurisdictions, the civil cases can be connected to a criminal case. The question now is if the negligent administrator can be charged for their negligence, which has caused damage and is the criminal responsibility justified in that situation.

As de Villiers has shown, the foreseeability of an attack is very important to see if the administrator is liable, such as for buffer overflow cases.[237] The foreseeability is very important in the healthcare sector cases which we looked before. Under the Health Insurance Portability and Accountability Act (HIPAA), healthcare centers must take measures that keep patient health data secure. If patient health data is leaked then the healthcare center may be liable because of HIPAA violation,[238]as

[231] De Villiers (2005), *supra nota* 33, 15.
[232] *Ibid.*, 15-16.
[233] *Ibid.*, 16.
[234] *Ibid.*
[235] *Ibid.*, 16, 29, 30.
[236] *Ibid.*, 17.
[237] *Ibid.*, 52.
[238] Harkins (2018), *supra nota* 67, 160.

well as because the protection of personal data is a fundamental right in accordance with the European Union Functioning Treaty Article 16(1) and Article 8 of the Charter on Fundamental Rights of the European Union.

But at the same time, it seems that the FDA regulations consider software updates in medical devices to be „a change in the medical device, and therefore must be evaluated to determine what obligations the manufacturer has under FDA regulations.“[239] Williams writes that FDA regulations can be an obstacle to update development, but software updates can also cause device malfunctioning, which means that the updates must be checked before putting them on the market.[240] Williams argues that this means that FDA approval is not needed for updates made for cybersecurity purposes, but they need to be tested to make sure that they do not change the device's function.[241]

2.4. Service provider liability

One of the most important barriers between malware and a potential victim is network security. The more secure a network is, the more difficult it is to penetrate the network.

De Villiers has shown that the EFR doctrine which I mentioned earlier has been used against ISPs who allowed free radicals to infringe copyrights by giving them infrastructure to do this.[242] This has been suggested as a method to fight online malware.[243] In the U.S, ISPs are protected against copyright infringement happening through their network by four protective parts of the law.[244] Copyright liability limitation for service providers excludes liability for automated transmission of data through their network if the user initiated and directed it and if the material is not

[239] Williams, K. Updates are Not Available: FDA Regulations Deter Manufacturers from Quickly and Effectively Responding to Software Problems Rendering Medical Devices Vulnerable to Malware and Cybersecurity Threats. Wake Forest J. Bus.& Intell. Prop. L.,14(3), 2014, p 370. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/wakfinp14&i=373, 25 March 2020

[240] *Ibid.*, 371.

[241] *Ibid.*, 401.

[242] De Villiers (2005), *supra nota* 33, 16.

[243] *Ibid.*, 17.

[244] Kao, A., RIAA v. Verizon: Applying the Subpoena Provision of the DMCA. *Berkeley Tech. L.J.*, 19, 2004, p 409. Retrieved from: https://btlj.org/data/articles2015/vol19/19_1_AR/19-berkeley-tech-l-j-0405-0426.pdf, 25 March 2020

modified.[245] The service providers are also protected when they do system caching.[246] If infringing material is stored on service provider's system or network and service provider has no knowledge or control of the infringement, there is no liability.[247] The service provider also has no liability if they provide a link to a infringing website and they are not aware of the infringement.[248]

When ISPs provide services, they usually act as intermediates who passively transfer data from one user to others. Usually, this process is automated, and the service provider has no control over the nature of the data or its destination. ISPs and web hosting services are normally free from liability that concerns user content on their website.[249] This has been confirmed when user-created malware was sent through the service provider's network without the service provider's knowledge.[250]

Talking about vulnerabilities in the network, the same principles apply to ISPs like they apply to any system or network administrator. There is a duty of care which the service provider must follow. This includes patching found vulnerabilities and actively trying to make the threat of a breach smaller. If the ISP fails to take preventive measures, then it will develop liability issues. Especially when the attack is foreseeable, and this is likely to occur through the existing vulnerability. Verizon has been criticized because it did not take measures to close their network vulnerability and this resulted in a worm attack.[251] To determine the extent of the ISP liability, the harm that is done plays quite a big role too – bigger harm means higher liability.[252]

However, not only network service providers need to make their network reasonably secure. Web browsers are used to access websites and because of that function, almost all users who have Internet access use one or more browsers. But browsers can have serious flaws in them that make users devices vulnerable to attacks.[253]

[245] 17 U.S.C. Ch. 5 § 512(a). Retrived from: https://www.govinfo.gov/content/pkg/USCODE-2011-title17/pdf/USCODE-2011-title17-chap5-sec512.pdf, 7 May 2020.

[246] 17 U.S.C. Ch. 5 § 512(b). Retrived from: https://www.govinfo.gov/content/pkg/USCODE-2011-title17/pdf/USCODE-2011-title17-chap5-sec512.pdf, 7 May 2020.

[247] 17 U.S.C. Ch. 5 § 512(c). Retrived from: https://www.govinfo.gov/content/pkg/USCODE-2011-title17/pdf/USCODE-2011-title17-chap5-sec512.pdf, 7 May 2020.

[248] 17 U.S.C. Ch. 5 § 512(d). Retrived from: https://www.govinfo.gov/content/pkg/USCODE-2011-title17/pdf/USCODE-2011-title17-chap5-sec512.pdf, 7 May 2020.

[249] Deutsch, J., Garrie, D. Instegogram: A New Threat and Its Limits for Liability. *J.L. & Cyber Warfare*, 6, 2017, pp 3-4. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/jlacybrwa6&i=3, 25 March 2020

[250] *Ibid*., 5-6.

[251] De Villiers (2005), *supra nota* 33, 53-54.

[252] *Ibid*., 54-55.

[253] Clarke (2007), *supra nota* 111, 20-23.

3. CHALLENGES AND RECOMMENDATIONS

3.1. Proposal for a liability system for AI based malware

Legislative changes are needed because criminal and civil law should ensure harmonization and should also contribute to public oversight and security. Also, in addition to the RoboLaw project, the heads of Member-States should launch more projects that can review the existing legal framework and adapt it to the new innovations and the risks associated with them.[254]

But the discussion in 2.1 showed that it is complicated to define a specific person, as responsible for damages caused by AI written malware, due to the unpredictability of AI systems and the ability to develop and continuously to change. Another serious problem arises here, since the institutions that control AI development and which are working on a new legal framework for the regulations and functioning of robotic technologies must consider the fact that law must be formal, consistent and stable since in all common law systems must be stable, which is supported by judicial precedents. However, AI and in particular, AI written malware is an unexplored area of judicial practice.

And there is a legal gap, which complicates the effective work of the judiciary in this area, because in all cases when the programmer is considered liable, there will be disagreement between the parties, because, if the software was developed using open source code or there were many different specialists who developed the program, then it will be complicated to determine a creator or the person responsible for the possible fault or purposeful malware. And if to consider the situation where AI has consciousness and is capable of self-development and autonomy, it will become even more complicated for legislation, because all inventions created by humans cannot be held responsible for the damage caused, because they are perceived as objects of property.[255] Also, it is not clear who will be liable for the damage if the programmer's computer is hacked, as in the case of *R v Green* [256], and an attack will be carried out on his behalf. Therefore, the author concludes that all participants should be liable for any malpractice and damage in case of illegal activity according to their contribution to the creation of the program or robotics.

[254] Pagallo (2013) *supra nota* 116, 98.
[255] Radutniy, O. E. Criminal liability of the artificial intelligence. *Problems of Legality*, no. 138, 2017, pp. 136. Retrivered from http://plaw.nlu.edu.ua/article/view/105661/106117, 27 Jan 2020.
[256] Exeter Crown Court, Unreported, *R v Green*, October 2003

However, in case, when it is difficult to identify the culprit or a liable person for the damages, it is reasonable to demand compensation from a natural person or a legal person who gets some benefits from the creation of technology or computer programs, and which is financially stable and able to cover the costs of damage[257].

AI can be seen as a severe source of danger, and a programmer or a manufacturer on whose behalf it acts should be held accountable. Given AI's widespread uses, it would be useful to implement the theory of "deep pocket" to update the General Product Safety Directive 2001/95/EC[258] and the Product Liability Directive 85/374/EEC[259] to ensure a software manufacturer's liability. Therefore, it means that a person engaged in harmful activities but gets some benefits from them must compensate for the caused damages to affected people. Thus, a person with a "deep pocket", whether it is a legal person or a programmer, is obliged to ensure against civil liability as a guarantee of his dangerous activity. Then, compensation should be proportionate to the likely damage that the malware will cause.

All Member-States need to develop a common approach to the constituent elements of criminal and civil offenses based on precedents related to illegal cyber-attacks, unlawful interference with data, and unlawful interception. National laws should take into account various circumstances during legal proceedings: a place of work and activity of the perpetrator, in particular, to make sure if she or he has knowledge and access to information systems. Therefore, in my opinion, it would be reasonable to improve the legislation and the Directive 2013/40/EU[260], since there are many sanctions and penalties for legal entities, but not for individuals and states.

Modern technologies and gadgets are developing rapidly, which does not allow them to be adequately tested, which means cyber-attacks are inevitable. AI written malware can have both local and transnational dimension, which emphasizes "the need for further action to approximate criminal law in this area" [261] as stated in the Directive 2013/40/EU.

Despite the fact that Member-States cooperate well with each other and the Union regarding the security of information systems and computer data. On the example of *Regina v Aaron Caffrey*[262]

[257] Čerka (2015), *supra nota* 27, 387

[258] OJ L 11, 15.01.2002

[259] OJ L 210, 07.08.1985

[260] OJ L 218, 14.8.2013

[261] *Ibid*.

[262] Southwark Crown Court, Unreported, *Regina v Aaron Caffrey,* 17.10.2003

case, we could see how vital professional specialists are. Also, it is advisable to instruct the jury to better understand the problems associated with spyware, as well as to know how the computer works as a whole system or not to convene the jurors at all.

CONCLUSION

There are many dangers associated with the use of AI, and one of the capabilities of AI is the creation of various malicious codes, and such codes that develop themselves. However, which actors would bear liability for damages caused by malicious code generated by AI?

The aim of this paper was to examine and to determine the limits of criminal and civil responsibility for AI-created malware between the creator and trainer of AI model in different scenarios, including intentional creation, accidental foreseeable, and accidental unforeseeable creation. Thus, understanding the nature and scope of obligations to malware created by AI between the creator and trainer of the AI model is an important step in helping states to meet their duties.

There is a liability problem for AI-written malware in criminal and civil law. Many actors may be involved in the creation and distribution of malware, even without knowing it. The transnational use of AI in the modern world emphasizes that malware can become a real threat, both for corporations and individuals, and for society and states. Therefore, it is necessary to correctly identify the participants who worked on the creation and distribution of malware, to limit possible damages as well as to prosecute the perpetrators.

Cyber-attacks and errors in the system are fraught with theft of confidential patient data, as well as personal data of clients of financial entities. The consequences of such malware can be very devastating, both for the reputation and image of companies and hospitals, and for patients and clients themselves.

At the moment, we know that in countries with traditional civil law, the damage should be compensated by the person responsible for the damage. Based on this law, it can be concluded that the developer is the responsible person who will be obliged to compensate for the damage. Still, many factors can affect the identification of the responsible person because malware is spreading transnationally.

Perhaps the most effective responsibility is related to the possible consequences of using AI and should be divided into two types. When a programmer gives the AI the final goal and write it in the original algorithm for certain violations, but the system exceeded its authority and committed a crime without the knowledge of the programmer. The second type concerns cases when the programmer did not program the AI system to commit a crime, but it happened. Then in the first planned crime, criminal liability will be divided into planned violations and unplanned violations. If the programmer planned that the system would commit certain violations, he or she would be fully liable for this crime. However, if, besides this crime, some other unforeseen violations occurred, in addition to the criminal liability for the planned crime, the programmer will be responsible for unplanned crimes.

An unplanned attack by a programmer will be considered as an accident. Still, only if he can prove it, then a programmer will be liable only for possible consequences and damage. This will be considered as mere negligence and will be seen under civil law. However, if the programmer knew about the error, but did not try to fix it and was inactive, then he or she will be responsible for the omission. However, if the programmer is not aware of the problems, he will not be held liable.

But in terms of the trainer's liability, if the creation of malware is not provided, then the trainer should not be held responsible, because he or she did not show negligence and did not foresee this situation. Also, the trainer may not consider an error serious. If the trainer uses computer software as input, which may include malware, then, in this case, he or she will be liable. The trainer can also take part in the development of new tools for the cybersecurity or forensic industry for which he or she will use malware. Training is necessary for all AI systems, because the machine must identify all available parameters and have a result, and subsequently develop its own algorithms to achieve the desired result. However, in this case, inadvertent improper training can lead to negligence. But laws can define the trainer's liability in different ways, so it all depends on national law and legal systems.

However, malware can also serve positive purposes. The AI system can be trained and then used to detect cyber-attacks based on data from the corresponding network or system. Corporations use similar methods to track the activities of their employees.

But, if a faulty program causes damage, there should be liability for the product, including liability for damage caused by defective products. If such a defect is found in the software or its damage, the developer’s negligence will be established, but not criminal liability. If an employee creates malware using AI, then this employee is fully responsible, but the company may also be partially liable because it does not have appropriate internal security measures. However, it may happen that an employee may accidentally instruct AI to create malware, in which case it will be considered as negligence in actions. However, the company still contributed to the creation of malware by providing favorable conditions.

The Administrator who created the vulnerability by negligence, which will be exploited by the attacker, will be liable for violation and damage under the EFR doctrine. However, this liability will be terminated if a deliberate attack occurs by using this vulnerability.

The consumer or a physical person can also be held liable when he or she downloads or installs unfamiliar programs and podcasts on the device. As a result, they are malware. Since knowledge of the possible risks and threats, but not the adoption of any measures or improvements to security measures, may lead to partial liability for damage. However, to establish legal liability, you must make sure that the person has an obligation, and this obligation has been violated. Moreover, if the computer has to process a large amount of data, this will complicate the process of making AI decisions. In this case, no one can be held responsible for the malfunctions, since it will not be possible to establish who is liable for AI-created malware.

AI can be seen as a severe source of danger; therefore, a programmer or a manufacturer on whose behalf it acts should be held liable. Thus, it would be useful to use the theory of "deep pocket" and update the General Product Safety Directive 2001/95/EC and the Product Liability Directive 85/374/EEC to ensure a software manufacturer’s liability. Therefore, it means that a person who is engaged in harmful activities but earns money or gets some benefits must compensate for the caused damages to affected people from the received profit.

Given the lack of regulations at the national level related to the prosecution of AI-created malware and perpetrators who instruct AI to generate malicious codes, the Member-States need to adhere to the same constituent elements of criminal and civil offenses based on precedents related to cyber-attacks. National laws must review various circumstances during the legal process, especially the background of the person who committed the crime, in particular, to make sure if she or he had access to information systems. Therefore, in my opinion, it would be reasonable to improve the national legislation and the Directive 2013/40/EU, since there are many sanctions and penalties for legal entities, but not for individuals and states.

REFERENCES

Books

1. Hallevy, G.,. (2015). *Liability for Crimes Involving Artificial Intelligence Systems*, Switzerland: Springer.
2. Joshi, K. B. (2010). *Data Structured and Algorithms in C++*. New Dehli: Tata McGraw Hill Education Private Limited.
3. Pagallo, U. (2013). *The laws of robots: crimes, contracts, and torts*. Law, Governance and Technology Series,10, Netherlands: Springer Science & Business Media.
4. Shinder, D. L., Ed Tittel, (2002), *Scene of the Cybercrime. Computer Forensics Handbook*. (1st ed). Chapter 6. Understanding Network Intrusions and Attacks,. Syngress Publishing.

Articles

5. Arivudainambi, D., *et al.* (2019). Malware traffic classification using principal component analysis and artificial neural network for extreme surveillance. *Computer Communications*, 147, pp 50-57.
6. Athanasiadou, E., *et al.*, (2018). Camera Recognition with Deep Learning. *Forensic Sci. Res.*, 3(3), pp 210-218. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/forsr3&i=212, 25 March 2020
7. Bedeli, M., *et al.* (2018). Clothing Identification via Deep Learning: Forensic Applications. *Forensic Sci. Res.*, 3(3), pp 219-229. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/forsr3&i=221, 25 March 2020
8. Brenner, W.S., Carrier B., Henninger J. (2004). The Trojan Horse Defense in Cybercrime Cases, *Santa Clara High Tech. L.J.,* 21(1), pp 6-51. Retrieved from: http://digitalcommons.law.scu.edu/chtlj/vol21/iss1/1, 8 May 2020
9. Clarke, R., Maurushat, A. (2007). Passing the Buck: Who Will Bear the Financial Transaction Losses from Consumer Device Insecurity? *J.L. Inf. & Sci.*, 18(1), pp 8-56. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/jlinfos18&i=8 , 25 March 2020
10. Cofone, I. N. (2019). Algorithmic Discrimination is an Information Problem. *Hastings L.J.*, 70, pp 1389-1443.
11. Caviglione, L., *et al.* (2016). Seeing the Unseen: Revealing Mobile Malware Hidden Communications via Energy Consumption and Artificial Intelligence. *IEEE Transactions on Information Forensics and Security*, 11(4), pp 799-810.
12. Čerka, P., Grigienė, J., Sirbikytė, G. (2015). Liability for damages caused by artificial intelligence. *Computer Law & Security Review: The International Journal of Technology Law and Practice*, 31(3), 376-389.
13. Deutsch, J., Garrie, D., (2017). Instegogram: A New Threat and Its Limits for Liability. *J.L. & Cyber Warfare*, 6, pp 1-7.Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/jlacybrwa6&i=3, 25 March 2020
14. Davis, J. P. (2018), Law without Mind: AI, Ethics, and Jurisprudence. *Cal. W. L. Rev.*, 55(1), pp 165-220. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/cwlr55&i=175, 25 March 2020
15. De Villiers, M. (2005). Free Radicals in Cyberspace – Complex Liability Issues in Information Warfare. *Nw. J. Tech. & Intell. Prop.*, 4(1), pp 13-60. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/nwteintp4&i=17, 25 March 2020

16. Gerstner M.E. (1993). Comment, Liability Issues with Artificial Intelligence Software, 33(1), *Santa Clara Law Review,* p. 239 Retrieved from: http://digitalcommons.law.scu.edu/lawreview/vol33/iss1/7, 25 March 2020

17. Giuffrida, I. (2019). Liability for AI Decision-Making: Some Legal and Ethical Considerations. *Fordham L. Rev.*, 88, pp 439-456.
18. Hansen R. L. (1989-1990). The Computer Virus Eradication Act of 1989: The War against Computer Crime Continues. *Software L.J.*, 3, pp 717-754. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/softljou3&i=717, 25 March 2020
19. Harkins, M., Freed, A. M. (2018). The Ransomware Assault on the Healthcare Sector. *J.L. & Cyber Warfare*, 6(2), pp 148-164. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/jlacybrwa6&i=339, 25 March 2020
20. Kirby, C. A., (2006), Defining Abusive Software to protect Computer Users from the Threat of Spyware. *Computer L. Rev. & Tech. J.*, 10, pp 287-324, Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/comlrtj10&i=287 , 25 March 2020
21. Kirtley, J. E., Memmel, S. (2018). Rewriting the Book of the Machine: Regulatory and Liability Issues for the Internet of Things. *Minn. J.L. Sci. & Tech.*, 19(2), pp 455-514. Retrieved from: https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1448&context=mjlst 7 May 2020
22. Kemp, R.. (2018). Autonomous Vehicles – Who Will Be Liable for Accidents? *Digital Evidence & Elec. Signature L. Rev.*, 15, pp 33-47. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/digiteeslr15&i=37, 15 Feb 2020.
23. Kingston, J. (2018) Artificial Intelligence and Legal Liability. *Australian Product Liability Reporter,* 28 (3), pp269-279. Retrieved from https://arxiv.org/pdf/1802.07782.pdf, 8 April 2020
24. Kroczynski, R. J. (2008). Are the Current Computer Crime Laws Sufficient or Should the Writing of Virus Code Be Prohibited? *Fordham Intell. Prop. Media & Ent. L.J.*, 18(3), pp 817-866. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/frdipm18&i=825, 25 March 2020
25. Kao, A. (2004). RIAA v. Verizon: Applying the Subpoena Provision of the DMCA. *Berkeley Tech. L.J.*, 19, pp 405-426. Retrieved from: https://btlj.org/data/articles2015/vol19/19_1_AR/19-berkeley-tech-l-j-0405-0426.pdf, 25 March 2020
26. Li, J.-H., (2018). Cyber security meets artificial intelligence: a survey. *Front. Inform. Technol. Electron. Eng.*, 19(12), pp 1462-1474.
27. Martinez, R.. (2019). Artificial Intelligence: Distinguishing Between Types & Definitions. *Nev. L.J.*, 19, pp 1025-1041.
28. McCarthy, J. (2007). What is artificial intelligence*? Stanford University, Computer Science Department*. p15. In: Čerka, P., Grigienė, J., Sirbikytė, G. (2015). Liability for damages caused by artificial intelligence. *Computer Law & Security Review: The International Journal of Technology Law and Practice*, 31(3), p378
29. Ningsih, A. S. (2019). The Doctrine of Product Liability and Negligence Cannot Be Applied to Malware-Embedded Software. *J Indonesian Leg. Studies*, 4(1), pp 7-20. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/jils4&i=11, 8 April 2020

30. O'Donnell, R. M. (2019). Challenging Racist Predictive Policing Algorithms Under the Equal Protection Clause. *N.Y.U. L. Rev.*, 94, pp 544-580.
31. Rodriguez, M. (2016). All Your IP Are Belong to Us: An Analysis of Intellectual Property Rights As Applied to Malware. *Tex. A&M L. Rev.*, 3(3), pp 663-690.. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/twlram2015&i=691, 25 March 2020
32. Semmler, S., Rose, Z. (2017-2018). Artificial Intelligence: Application Today and Implications Tomorrow. *Duke L. Rev.*, 16, pp 85-99. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/dltr16&i=85, 25 March 2020
33. Sloan, R. H., Warner, R.. (2018). Algorithms and Human Freedom. *Santa Clara High Tech. L.J.*, 35, pp 22-24.

34. Schuster, W., M.. (2018). Artificial Intelligence and Patent Ownership. *Wash. & Lee L. Rev.*, 75(4), pp 1945-2004. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/waslee75&i=1991, 25 March 2020
35. Scharre, P. (2019). Killer Apps: The Real Dangers of an AI Arms Race. *Foreign Aff.*, 98(3), pp 135-144. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/fora98&i=557, 25 March 2020
36. Swanson, G. (2019). Non-Autonomous Artificial Intelligence Programs and Products Liability: How New AI Products Challenge Existing Liability Models and Pose New Financial Burdens. *Seattle U. L. Rev.*, 42, pp 1201-1222.
37. Sherman, C. R. (1998). The Surge of Artificial Intelligence: Time To Re-examine Ourselves. Implications: Why Create Artificial Intelligence? In: Čerka, P., Grigienė, J., Sirbikytė, G. (2015). Liability for damages caused by artificial intelligence. *Computer Law & Security Review: The International Journal of Technology Law and Practice*, 31(3), p377
38. Terry, N. P. (2018). Appification, AI, and Healthcare's New Iron Triangle. *J. Health Care L. & Pol'y*, 20(2), pp 117-182, Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/hclwpo20&i=127, 25 March 2020
39. Tuthill, G.S. (1991). Legal Liabilities and Expert Systems, *AI Expert.* In Kingston, J. (2018) Artificial Intelligence and Legal Liability. *Australian Product Liability Reporter,* 28 (3), p269-279. Retrieved from https://arxiv.org/pdf/1802.07782.pdf ,25 March 2020
40. Valentine, S. (2019). Impoverished Algorithms: Misguided Governments, Flawed Technologies, and Social Control. *Fordham Urb. L.J.*, 46, pp 364-427.
41. Violago, V., Quevada, N. (2018). AI: The Issue of Bias. *Managing Intell. Prop.*, 227, pp 32-36. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/manintpr277&i=34, 25 March 2020
42. Wendel, B. W. (2019). The Promise and Limitations of Artificial Intelligence in the Practice of Law. *Okla. L. Rev.*, 72, pp 21-49
43. Weyhofen, C. (2019). Scaling the Meta-Mountain: Deep Reinforcement Learning Algorithms and the Computer-Authorship Debate. *UMKC L. Rev.*, 87(4), pp 979-996, Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/umkc87&i=1013, 25 March 2020
44. Williams, K. (2014). Updates are Not Available: FDA Regulations Deter Manufacturers from Quickly and Effectively Responding to Software Problems Rendering Medical Devices Vulnerable to Malware and Cybersecurity Threats. *Wake Forest J. Bus. & Intell. Prop. L.*, 14(3), pp 367-417. Retrieved from: https://heinonline.org/HOL/P?h=hein.journals/wakfinp14&i=373, 25 March 2020

Estonian legislation

1. KarS RT I, 28.02.2020, 5, §216(1)(1)
2. KarS RT I, 28.02.2020, 5, §217(1)

EU and international legislation

1. Consolidated version of the Treaty on European Union and the Treaty on the Functioning of the European Union 2012/C 326/01. Official Journal C 326, 26.10.2012
1. Charter on Fundamental Rights of the European Union 2012/C 326/02. Official Journal C 326, 26.10.2012
2. Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products. Official Journal of the European Union, L 210, 7 Aug 1985.

3. Council Directive 2001/95/EC of 3 December 2001 on general product safety. Official Journal of the European Union, L 11, 15 Jan 2002
4. Directive 2013/40/EU of the European Parliament and of the Council of 12 August 2013 on attacks against information systems and replacing Council Framework Decision 2005/222/JHA
5. Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe, Brussels, 25.4.2018 COM(2018) 237 final.
6. U.S. Code - Unannotated Title 18. Crimes and Criminal Procedure §1030. Fraud and related activity in connection with computers Retrieved from: https://www.govinfo.gov/content/pkg/USCODE-2010-title18/pdf/USCODE-2010-title18-partI-chap47-sec1030.pdf, 7 May 2020

Court decisions

1. *U.S. v. Morris,* 928 F.2d 504 (2d Cir. 1991), Mar. 7, 1991. Retrieved from https://casetext.com/case/us-v-morris-123
2. Southwark Crown Court, Unreported, Regina v Aaron Caffrey, 17.10.2003
3. Exeter Crown Court, Unreported, *R v Green*, October 2003

Other sources

1. Avgerinos, T., *et al.*, (2011), AEG: Automatic Exploit Generation. *NDSS Symposium 2011*, Conference paper. Retrivered from http://security.ece.cmu.edu/aeg/aeg-current.pdf, 27 Jan 2020.
2. Job Wizards. Do you know what algorithm trainers do and how cognitive systems work? Retrivered from https://job-wizards.com/en/cognitive-systems-what-do-algorithm-trainers-do/, 7 May 2020.
3. Pfeffer, A., et al., (2017), Artificial Intelligence Based Malware Analysis. *ArXiv*. Retrivered from http://arxiv-export-lb.library.cornell.edu/pdf/1704.08716, 27 Jan 2020.
4. Radutniy, O.E. 2017. Criminal liability of the artificial intelligence. *Problems of Legality*, no. 138, p. 136. Retrivered from http://plaw.nlu.edu.ua/article/view/105661/106117, 27 Jan 2020.

Appendix 1. Non-exclusive licence

Non-exclusive licence for reproduction and for granting public access to the graduation thesis[1]

I ____________________________________Valeria Gaiduk_______________________________
(author's name)

1.Give Tallinn University of Technology a permission (non-exclusive licence) to use free of charge my creation

__________________Legal Responsibility in the context of AI-written malware__________________
(title of the graduation thesis)

supervised by _________________________Agnes Kasper___________________________________
(supervisor's name)

1.1. to reproduce with the purpose of keeping and publishing electronically, including for the purpose of supplementing the digital collection of TalTech library until the copyright expires;

1.2. to make available to the public through the web environment of Tallinn University of Technology, including through the digital collection of TalTech library until the copyright expires.

2. I am aware that the author also retains the rights provided in Section 1.

3. I confirm that by granting the non-exclusive licence no infringement is committed to the third persons' intellectual property rights or to the rights arising from the personal data protection act and other legislation.

[1] The non-exclusive licence is not valid during the access restriction period with the exception of the right of the university to reproduce the graduation thesis only for the purposes of preservation.