SLCMC 2021
International conference «State and law in the context of modern challenges»

ARTIFICIAL INTELLIGENCE AND CRIMINAL LAW RISKS

Natalia A. Lopashenko (a)*, Elena V. Kobzeva (b), Zaur D. Rozhavskiy (c)
*Corresponding author

(a) Saratov State Law Academy, 1, Volskaya Str., Saratov, 410056, Russia, lopashenko@yandex.ru
(b) Saratov State Law Academy, 1, Volskaya Str., Saratov, 410056, Russia, evkobzeva@yandex.ru
(c) «Lamel Consult» LLC, 1, Kirova St., Astrakhan, Russia, llc.lamel@ya.ru

Abstract

The rapid scientific development leads to the outspread of digital technologies. However, the growing digitalization in the society, along with positive results, causes new unknown forms of socially dangerous behavior (activity), which consequently harms criminal protection objects. In many respects, such criminal law risks are due the phenomenon of artificial intelligence and its application in various spheres of human life. "Mainstream Science on Intelligence", widely recognized in the scientific community, defines the concept of intelligence as a general mental capacity, including the ability to conclude, plan, solve problems, think abstractly, understand complex ideas, learn quickly and get knowledge from experience. This interpretation of intelligence presupposes that the described mental processes become quiet expected from non-human intelligence, but it is too soon to equate artificial intelligence responsibility with the human one as Homo sapiens behavior is greatly influenced by feelings and emotions. Artificial intelligence maintenance can have a malicious impact on itself and (or) its holder that sometimes cannot be unambiguously qualified from the criminal law viewpoint. In these conditions, it is necessary to identify the existing criminal law risks and draw scientifically based conclusions for the further development of domestic criminal legislation. Obviously, its content will directly depend on the level of scientific and technological achievements in the artificial intelligence software engineering.

1. Introduction

The study of criminal legal personality and artificial intelligence is inextricably linked to utilitarianism and deontology, the ethical dilemmas of unmanned vehicles and cognitive science on the "trolley problem". The utilitarianism necessarily prescribes to switch the arrow so that the trolley goes on a different path to reduce the number of human victims, but this problem can be presented differently if to speak about the artificial intelligence. For example, a person will be tied to the rails on one of the tracks, but switching the arrow will entail the predicted destruction of the trolley itself due to rail defects on the second track (derailment) and thus the inability to complete the task of arriving at point "B".

Indeed, scientific progress cannot be stopped, but at present Russia is not among the leading countries in the development and implementation of computer technologies (with the exception, perhaps, at fiscal and financial sectors). And, therefore, the introduction of ethical and legal restrictions on scientific and technological development must be adequate to the existing risks. Otherwise, it may lead to a negative interest increase of the IT industry representatives in the domestic market as such.

2. Problem Statement

The problem of artificial intelligence legal personality is primarily due to the well-known cases of injuries caused by a robot to a person. The Internet presents enough examples of how a machine empowered with artificial intelligence caused the death of an individual who, in the opinion of such a machine, prevented the tasks performance facing it (Khizhniak, 2018). This clearly indicates that, despite the algorithms set in the development process, the man-made mind (due to the training or failures in it) gives priority to its own tasks implementation, and not the life of its creator.

In the course of artificial intelligence programming, developers are guided by the three laws of robotics, formulated by Isaac Asimov in his short story "Runaround":

1. A robot cannot harm a human being or, by its inaction, allow a human being to be harmed.

2. A robot must obey all orders given by a human being, except in cases where these orders are contrary to the First Law.

3. The robot must take care of its own safety to the extent that it does not contradict the First or Second Laws.

In order to provide legal support for scientific and technological progress which presupposes the artificial intelligence usage in various areas of human life, a number of countries have adopted regulations defining the general conceptual principles of civil law relations of a corresponding nature, as well as the requirements for the development and automatic software management. These include, for example: Decree of the Government of the Russian Federation No. 1632-r of July 28, 2017 "On the Program Approval "Digital Economy of the Russian Federation"; Decree of the President of the Russian Federation No. 203 of May 9, 2017 "On the Strategy for the Development of Information Society in the Russian Federation for 2017–2030"; National Robotics Initiative 2.0: Ubiquitous Collaborative Robots (USA); The Eighth Law on the Amendments to the Road Traffic Law of June 16, 2017 (Germany).

In this regard the question of whether criminal law should (or can) respond to the artificial intelligence introduction becomes very important (Kibalnik & Volosyuk, 2018).

3. Research Questions

There are certain problematic issues that humanity is already facing right now:

1) who should (if should at all) be held criminally liable for the harm caused to people's life and health in the functioning process of artificial intelligence carriers (drones, "smart" machines in complex industries)?

2) how to qualify a deliberate illegal impact on a mechanical carrier of artificial intelligence, for example, on a bionic prosthetic limb with artificial intelligence or an electrochemical eyeball embedded in the human body?

4. Purpose of the Study

The purpose of this work is to find scientifically based answers to the key questions of modern criminal law doctrine regarding artificial intelligence: in what capacity should artificial intelligence be considered an object of criminal law protection or a possible subject of criminal liability?

5. Research Methods

The methodological basis of the research is the universal dialectical cognitive method of studying phenomena and processes in the surrounding reality. A set of general scientific and private research methods (synthesis, induction, deduction, abstraction, system, structural-functional, logical, formal-legal, and comparative-legal analysis) has been also used in theoretical and practical work provisions.

6. Findings

According to many different opinions of major public figures, scientists, specialists in the field of programming, the rapid speed of artificial intelligence development, the growing geography of its use in various spheres of human life and the increasingly frequent incidents of its "disobedience" to a man, in the relatively near future, artificial intelligence can surpass human intelligence and, as a result, get out of the natural mind control. In the context of such forecasts, artificial intelligence is often compared to nuclear weapons that can endanger the very existence of humanity, destroy the human race as such. According to Bostrom (2014), in modern conditions, work on artificial intelligence proceeds in the direction where its "mental" potential is constantly increasing: from weak to strong. If this trend continues in the coming decades (approximately by the middle of the 21st century), ordinary artificial intelligence will be replaced with a high degree of probability by artificial intelligence of the human level, defined as "the ability to master most professions, at least those that an average person could own". This, in turn, is inevitable and quite quickly, according to the scientist, "explosively", will give rise to the creation of an artificial superintelligence, i.e. a mind that is not just equal to the human mind, but superior to it, and therefore capable of "leading to huge consequences – both extremely positive and extremely negative, up to the death of humanity" (Bostrom, 2014, pp. 12–15, 28–31).

Even ten or even seven years ago, in the year of N. Bostrom's book publication, most readers perceived his theory as a science fiction, but today, taking into account the achieved level of artificial

intelligence development, the concerns expressed by the scientist sound quite real, gaining more and more supporters. Of course, this cannot but involve the criminal law science, since the excesses of artificial intelligence that arise during its operation, contrary to the expectations of developers, often form similarities with individual crimes and encourage their criminal legal assessment. So, the story of launching the so-called chatbot with the female name Tai, designed to communicate with young people, received a great resonance on Twitter. During the day Tai, communicating with social network users, has been inspired by the ideas of misanthropy, racism, repeatedly speaking in support of Hitler's policy and genocide, as well as confessing hatred for feminists and other social groups. The developer of the program, Microsoft, urgently disabled the chatbot, apologized to the users, deleted all comments and started improving the algorithm of its work (Krivets, 2016).

Thus, it should be noted that the modern doctrine of criminal law is significantly late in developing rules for the criminal legal assessment of harm caused to the interests protected by criminal law in the process of using artificial intelligence. And the point is not even in the global negative consequences that may occur tomorrow and the day after tomorrow in connection with the appearance of "superintelligence", but in those negative and dangerous social results of its use that take place already today, and to be honest, happened even yesterday.

So, prior to the artificial intelligence introduction into everyday use, it is trained to achieve the main goal of its creation. At the same time, based on the laws of robotics, information technology experts precede this process by establishing a ban on program modification by itself in addition to or against the will of the creators (authors). This is a key requirement for the process of creating artificial intelligence, compliance with which will prevent or minimize earlier negative consequences, so the question of establishing a criminal legal obligation to fulfill it deserves its independent consideration in the future.

In the process of acquiring the ability to make autonomous decisions in conditions *where it is impossible to write and correct its program code independently*, artificial intelligence, based on the algorithms specified by the developer, analyzes, synthesizes, compares, induces, deducts and uses other cognitive methods related to the entire volume of incoming information, but is not able to perceive the decision made by it and (or) a complete/non-complete operation.

Thus, artificial intelligence at the current time does not have any mental processes; it has no ability to feel inner emotions for the external manifestations of its functionality. Therefore, it cannot be found guilty (in the criminal law aspect) of committing a crime, although it can act as a criminal behavior element on its formal grounds.

The foreign scientific literature proposes the following liability types for illegal acts connected with the artificial intelligence maintenance:

- strict liability of artificial intelligence;
- another person's liability when artificial intelligence becomes a crime tool;
- vicarious liability of the operator's management;
- person's liability for the improper conduct under a contract or tort involving natural and probable consequences of act or omission that directly caused the plaintiff's damage or harm (Hallevy, 2012).

It is obvious that hypothetical attempts to give artificial intelligence the status of an independent subject of criminal responsibility are broken down due to the subjective assessment of actor's performance accompanied by harm to the object of criminal protection. It is an axiom that machine intelligence is currently unable to have feelings as there is no experience in accumulating biochemical reactions, the nature and essence of which have not yet been sufficiently studied by mankind. At the same time, the key value will not be just the ability of the machine to experience feelings and emotions, but the capacity of unnatural intelligence to realize the nature and social danger degree of its behavior.

It seems that the thesis about the impossibility of recognizing artificial intelligence as the crime subject will remain relevant until a certain man-made artificial intelligence is declared to be capable of experiencing both biomechanical and biochemical reactions, which are subsequently classified as mental processes. Until this obstacle is removed, the developer (an individual, group of persons) or a commercial (non-commercial) structure) and the end user must be liable for offenses committed by a hardware and software artificial intelligence complex, depending on its extent of guilt (Joh, 2016; Pagallo, 2011; Van Riemsdijk et al., 2013).

This will require a complex computer-technical expertise that can determine the reason why artificial intelligence considered the operation to be permissible. In particular, if such a failure occurred by the errors in code algorithm (then the developer, treated as a legal entity, must be liable for the sale of goods (services) that do not meet the security requirements). Or this can be due to the intentional writing of "harmful" code or the deliberate harmful artificial intelligence training (then the developer, operator or an unauthorized intruder into the program should be held liable for intended willful crime against the individual, society or the state). It should be agreed that all cases when the interests protected by criminal law may be harmed during the operation of artificial intelligence and, as a result, there will be a need for a criminal legal response, it will be optimal to reduce to four situations: 1) an error was made in the process of artificial intelligence creation, which resulted in harm; 2) illegal access was made to the artificial intelligence program, which caused its damage or modification, as a result of which harm was caused; 3) artificial intelligence, which has the ability to self-study, made a decision to commit actions/inaction that caused harm; 4) artificial intelligence was created by criminals to cause harm (Khisamova & Begishev, 2019).

If a crime related to the artificial intelligence maintenance is committed with the direct intent, then it has quite certain prospects in terms of criminal law qualification; crimes committed with indirect intent or through negligence raise many questions. The latter circumstance is because of the uncertain social relations variety in which such a person can be involved. In addition, if to speak about the levity then the abstract foresight of socially dangerous consequences in artificial intelligence maintenance becomes almost illusory. And if we talk about negligence – then the establishment of duty and the possibility of foreseeing the corresponding results in most cases will have no necessary point of support. In this regard, the scientific thought focused on clarifying the criminal-legal content of guilt related to artificial intelligence maintenance is in great demand.

Among other things, the hypothesis about the possible criminal prosecution of artificial intelligence is disrupted by the high risks of its infection with malware in the absence of a panacea for this

threat (Brenner et al., 2004). In this case, the liability should be borne by the virus developer (Shestak et al., 2019).

At the same time, it seems that at present time, in the interests of ensuring generally recognized objects of criminal legal protection, there is a need to introduce legislative prohibitions on the malicious development and use of artificial intelligence, on harmful interaction with it, as well as other effects on artificial intelligence without separation from its physical carrier under the threat of criminal liability (taking into account the provisions on justified criminal legal risk).

Artificial intelligence can and should be the subject of criminal protection. Today – it is simply like a smart machine, a kind of physical shell, information program, and later – in any form and variants invented by mankind. In this case the artificial intelligence and social relations involving it must be under criminal protection from causing harm, destruction, liquidation but, apparently, only under the condition of its legal origin and existence of man-made intelligence.

7.   Conclusion

The modern potential of artificial intelligence excludes the question of its criminal legal personality. Although the socially dangerous activity of the carrier of non-human intelligence may take place, however, the ability to its internal (mental) perception, which is a prerequisite for criminal liability, is absent in artificial intelligence. The developer of the corresponding program or the operator of artificial intelligence device must be responsible for causing harm to the interests protected by criminal law in connection with artificial intelligence operation. To increase the task effectiveness in criminal law, it is necessary to introduce criminal liability of legal entities (developers of artificial intelligence) in Russia, and a criminal law ban on the creation of an artificial intelligence program with the possibility of its independent completion in addition to or against the will of the authors, as well as to clarify the intellectual and volitional content of guilt in the form of indirect intent, levity and negligence. The exceptional and completely unknown nature of the technical-biological (and other) coexistence of artificial intelligence and its carrier reveals the insufficiency of criminal law protection and the need to optimize it in terms of recognizing artificial intelligence as an independent object of such protection. It is also appropriate to support the thesis that the legal support of activities related to the creation and operation of artificial intelligence should be developed consistently and in a timely manner, taking into account all possible legal risks and in the direction of ensuring a balance between the interests of society and individuals (Ponkin & Redkina, 2018).

References

Bostrom, N. (2014). *Superintelligence Paths, Dangers, Strategies.* Oxford University Press.
Brenner, S. W., Carrier, B., & Henninger, J. (2004). The Trojan horse defense in cybercrime cases. *Santa Clara Computer & High Tech. LJ, 21*, 1.
Hallevy, G. (2012). Unmanned Vehicles – Subordination to Criminal Law under the Modern Concept of Criminal Liability. *Journal of Law, Information and Science, 21*, 200–211.
Joh, E. E. (2016). Policing police robots. *UCLA L. Rev. Discourse, 64*, 516.
Khisamova, Z. I., & Begishev, I. R. (2019). Criminal liability and artificial intelligence: theoretical and applied aspects. *All-Russian Journal of Criminology, 13*(4), 564–574.

Khizhniak, N. (2018). *HI-News 10 cases of robots that killed people.* https://hi-news.ru/robots/10-sluchaev-s-robotami-ubivshimi-lyudej.html#dzhoshua_braun

Kibalnik, A. G., & Volosyuk, P. V. (2018). Artificial intelligence: questions of criminal law doctrine awaiting answers. *Legal Science and Practice: Bulletin of the Nizhny Novgorod Academy of the Ministry of Internal Affairs of Russia, 4*(44), 173–178.

Krivets, A. (2016). *"The Mirror of Society". The story of Microsoft's misanthropic bot.* https://medialeaks.ru/2603nastia_tay/

Pagallo, U. (2011). Killers, Fridges, and Slaves: A Legal Journey in Robotics. *AI & Society, 26*(4), 347–354.

Ponkin, I. V., & Redkina, A. I. (2018). Artificial intelligence from the point of view of law. *Bulletin of the Peoples ' Friendship University of Russia. Series: Legal Sciences, 22*(1), 91–109.

Shestak, V. A., Volevodz, A. G., & Alizade, V. A. (2019). On the possibility of a doctrinal perception by the common Law system of artificial intelligence as a subject of crime: by the example of US criminal law. *All-Russian Journal of Criminology, 13-4*, 547–554.

Van Riemsdijk, M. B., Dennis, L. A., Fisher, M., & Hindriks, K. V. (2013). Agent reasoning for norm compliance: A semantic approach. *Proceedings of the 12th international conference on Autonomous Agents and Multi-Agent Systems*, 499–506.

.