Criminal Liability of the Artificial Intelligence

Danila Kirpichnikov¹*, Albert Pavlyuk², Yulia Grebneva³, Hilary Okagbue⁴

¹ Kazan Innovative University named after V. G. Timiryasov (IEML), 42 Moskovskaya str., 420111, Kazan, Russia
² Moscow State Institute of International Relations (MGIMO), 76, Prospect Vernadskogo, 119454, Moscow, Russia
³ Kuban State Agrarian University named after I.T. Trubilin, ul. Kalinina, 13, 350044, Krasnodar, Russia
⁴ Covenant University, Ota, Nigeria km. 10 Idiroko Road, Canaan Land, Ota, Ogun State, Nigeria

Abstract. Today, artificial intelligence (hereinafter – AI) becomes an integral part of almost all branches of science. The ability of AI to self-learning and self-development are properties that allow this new formation to compete with the human intelligence and perform actions that put it on a par with humans. In this regard, the author aims to determine whether it is possible to apply criminal liability to AI, since the latter is likely to be recognized as a subject of legal relations in the future. Based on a number of examinations and practical examples, the author makes the following conclusion: AI is fundamentally capable of being criminally liable; in addition, it is capable of correcting its own behavior under the influence of coercive measures.

1 Introduction

Rapidly developing public relations with the participation of artificial intelligence form significant criminological risks: information that is accessed in information and telecommunications networks due to its physical properties is exposed to the greatest threat of illegal influence coming from artificial intelligence, software and electronic information systems of critical information infrastructure objects currently do not have the means and methods of protection from attacks of such cybernetic entities, this significantly actualizes the development of science-based positions on the relative legal regulation of the position of artificial intelligence and the development of legal models of protection from attacks that can potentially come from it. In our research, we rely on the thematic base of national legislation (an example of this is the provisions of the acts of The government of the Russian Federation), scientific works of domestic and foreign experts published in different periods of time, combined research of artificial intelligence, General scientific works, as well as information distributed by the media.

It is now common knowledge that digital innovation can be used for both the public good and criminal purposes. And first of all, it concerns cybercrime, the level of aggression of which increases every year. It is for this reason why Europol and its partner organizations are fighting cybercriminals across the Board.

* Corresponding author: sauleat@yandex.ru
Today researchers note the following as a trend: the advance of legislative regulation of doctrinal developments. For this reason, criminal law science often criticizes the provisions of the criminal code consistently and more reasonably. It seems to be that such a situation is unacceptable. Legislative changes, and especially criminalization, should objectively foresee changes in the structure of a certain sphere of public relations and promptly respond with a solvent legislative position that in turn can be largely based on the approved theoretical developments of science [1]. Humanity is on the threshold of an era when the expansion of the horizons of AI application unleashes a new digital revolution. Among the areas that are characterized by a high level, the field of digital technology occupies one of the leading places. And there is one of the industries carrying the greatest criminological risks, the activity on creation, introduction and use of AI acts [2, 3]. As rightly noted by scientists, the pace of development of systems and devices with AI will cause the need for a total revision of all branches of law [3, 4]. Of course, the introduction of AI in human activities is extremely promising, because the amount of data accumulated and processed by it, clearly exceeds the existing systems today, involving human participation [5, 6].

Today there is a gradual introduction of intelligent technologies in many areas of human activity. Today there is a gradual introduction of intelligent technologies in many areas of human activity. So, for example, on September 27, 2019, the striking unmanned aerial vehicle "Hunter" made a flight in an automated mode in full configuration with access to the airborne alert area [7, 8]. Currently developing software that can recognize human emotions on eye movement and facial muscles and a learning algorithm, equipped with a biometric sensor capable of providing medical services are much superior to those now available in the best medical organizations in the world [9, 10].

In turn, the resolution of the European economic and social Committee of 31 August 2017 "Artificial intelligence-the consequences of AI in the (digital) single market, in production, consumption, employment and society" indicates that the introduction of AI in social practices can improve the effectiveness of activities in the implementation of many goals, namely-the eradication of poverty, transportation safety, quality medicine, more personalized education, industry development [11, 12]. We will demonstrate on the following table the range of spheres of public life predicted by scientists in which, according to expert estimates expressed in material equivalent, artificial intelligence will become the most widespread (Figure 1).

![Image of a table showing the forecast of the world artificial intelligence market volumes up to 2025 expressed in billions of US dollars.](image)

**Fig. 1.** The range of spheres of public life predicted by scientists, in which, according to expert estimates expressed in material equivalent, artificial intelligence will become the most common.
2 Methodology

In the course of this research, we used: a system-structural method, expressed in the allocation of their structure in integral systems of social relations with the definition of those social practices in which artificial intelligence is most likely to become widespread; a dialectical method of cognition, since the laws of materialistic dialectics are of universal importance and are equally characteristic of the development and functioning of artificial intelligence technologies; highly relevant to the purposes of the study a method of modeling, which consists in creating a mental model allows to obtain the expected information required for the disclosure of the main results of the study; the wide distribution of the received prediction method based on the analysis of the objective laws of development of artificial intelligence as a social and technical phenomena, and using the theory of prognostics in the development of further research; the analysis of interrelated provisions of scientific papers on similar or overlapping topics, and the synthesis of General theoretical developments for the purpose of learning artificial intelligence were also involved. The wide application of various methods in the work allows us to reasonably assume that the content will be relevant to the scientific community, and the conclusions will allow us to develop the doctrine of the legal regulation of artificial intelligence.

3 Content of the study

For a substantive understanding of the essence of this work, we should refer to the description of AI as a phenomenon of digital reality. According to one of the positions reflected in the literature, it is an artificial neural system that simulates human thought processes using the computing power of a computer. From another point of view, AI can be understood as the process of creating machines that will act in a way that will be perceived as intelligent by man [13]. It seems that there is no fundamental difference in the existing definitions, but each of them contains an indication of the property of reasonableness, which is relevant from the standpoint of criminal law. Intelligence as a sign of activity gives rise to AI ability to self-learning, in other words, initiates the ability on the basis of collecting, perception and processing of new information, to acquire properties and expand functionality beyond that provided by developers in the design and creation [14].

The process of AI decision-making is described in different ways in the literature, for example, Ushakov E. M. points out that AI consists of three layers of artificial neurons: the first receives input data of the surrounding reality, in the inner layer they are processed and transferred to the output layer, in which the result is formulated [7, 15].

A slightly different description is given in the work of M. Tim Jones, who, describing the decision-making process, based on the method of simulation of recovery. The researcher divides the decision-making algorithm into five stages, in which the AI develops several alternative possible solutions and, bases on the evaluation of the effectiveness of each of them, formulates a conclusion about the final decision. At the same time, it should be noted that the technical literature does not contain indications of human participation or involvement in AI activities [16-18].

Developing the thesis about the autonomy of AI in decision-making, we believe that the opinions expressed in the literature about the advisability of bringing to criminal responsibility for the actions of AI its developers [2] cannot be recognized as legitimate and justified. It is hardly possible to establish a direct causal relationship between the actions of one Autonomous entity (developer) and the second (AI). An exception may be the theory of
equivalence, which recognizes the cause of the consequences of any condition necessary for its occurrence.

According to the logic of those researchers (Denisov N. L.) [19], who suppose the criminal responsibility of developer for acts committed by AI, possible, prosecution for the death of the person who caused light injury to the victim, who on the way to the hospital got hit by a car, can be justified. In the affected context, it seems appropriate to give one of the definitions of cause-and-effect relationship to ensure the constructiveness and validity of further reasoning: causation - one of the varieties of determinative relationship, a relationship between phenomena, when one or more interacting phenomena (cause) naturally generate another phenomenon (effect).

With regard to the above, it is necessary to make a clarification: we use the term AI it would be correct to call a Strong AI. The separation in this case is fundamental, since all other AI, including electronic systems, with which a person comes into contact every day, are weak [3]. The reason for this is that they have only the ability to perform the only function for which they were created, and self-education, which entails an increase in knowledge and, accordingly, the expansion of functionality, is unusual for them. That is why it is not appropriate to consider them as a potential subject of harm.

It is also necessary to distinguish AI (hereinafter referred to as the abbreviation will be referred to only AI related to the category of strong) from Expert systems (ES-systems that are based on information bases to help in decision-making) [10]. First of all, strong AI is distinguished to have the ability to self-study and consciously volitional behavior, for this reason, this kind of AI is considered as a subject of criminal responsibility and is taken as the basis for further reasoning.

It seems that successful regulation of relations involving the AI should formulate a clear definition, here it is important to make it possible not only to distinguish a learning algorithm from the other phenomena of digital reality, but it also allows to differentiate AI from other algorithms of decision-making, or obtaining information that does not have the ability to self-learning and non-autonomy of the will.

The doctrine of criminal law, as well as other technical Sciences have developed the following definitions, which, as it seems to us, are important for reasoning about the subject of criminal responsibility. In particular, some researchers believe that AI is a science and technology that includes a set of tools that allow a computer to present answers to questions on the basis of accumulated knowledge and formulate expert opinions based on them, that is, to obtain knowledge that is not invested in it by developers [20, 21].

Other experts believe that AI is a mathematical model capable of learning, created in the likeness of the human brain [2]. Without denying the practical significance of the formulated definitions, we note that the above mentioned is a doctrinal development.

In terms of the legal consolidation in the regulatory system of the Russian Federation, as N. Kulikov correctly points out, the terms potentially capable of having relation to AI - "robot" and "robotics" appear only in a few legislative acts of far from paramount importance [3], with the exception of several state program-strategic legal acts 13. In the context of the above, perfectly logical statement I. R. Begishev that the resolution of the question of the legal fixing of the concept of AI should be to determine how the current legislation considers the possibility of the existence of such systems and can be used to organize relationships in terms of their use.

Of course, the introduction and application of AI will significantly change the really emerging social relations, so we believe that for a complete and comprehensive concept of AI, it is necessary to include a number of common features that, firstly, are repeated in each individual AI, secondly, are essential and thirdly, characterize the social significance of AI as a phenomenon. However, the formulation of the definition is an intermediate, although
necessary, task of forming the concept of criminal responsibility of AI as a necessary measure of legislative response to changes in modern social practices involving its participation [22].

We believe that these processes will entail a significant change in the structure of public relations, shifting the vector of crime in the direction of increasing the qualitative and quantitative use of machine learning technologies to commit crimes. However, this technology is able to act not only as a means of committing a crime, fixed as a feature that describes the objective side of the crime of a certain type.

Having the ability to consciously-volitional behavior and autonomy in decision-making, AI is able to commit an act containing signs of a crime of a certain type, being the subject of a crime, while being completely independent both in decision-making and in the completeness of the implementation of the objective side of the composition [19].

Modern criminal law leaves no other possibility than to identify an individual involved in any form of activity and the creation of AI. We believe that it is impossible to agree with anthropocentric positions of realization of criminal responsibility, because as mentioned above, due to the essential properties of AI, they contradict both the principle of guilt and justice. We believe that it is unacceptable to postpone legislative changes until the actual tendentiousness of committing AI crimes appears, since in this case criminalization will be untimely, accordingly, it is necessary to develop legal changes in the shortest possible time [3].

This thesis is confirmed, in particular, by the fact that for that time in the social practice cases of AI crimes are already known. So, in March 2018, the car of the Uber organization under the control of AI, as a result of equipment failure, hit a woman moving perpendicular to the roadway, which caused death. It was found that the AI mistakenly regarded it as "false operation" [10, 23].

On the basis of the above, it seems relevant and expedient to bring theoretical, doctrinal justifications of the possibility of recognizing AI as a subject of criminal responsibility. According to P.M. Morkhat [24], the resolution of the issue of AI as a subject of criminal responsibility is directly related to the possibility of conferring its rights and imposing duties. In this case, there are two alternatives:

- empowering of AI as a separate information entity, which finds external expression in certain units (devices, systems, robotic-automatic and software-hardware complexes), legal personality;
- empowering AI legal personality similar to the scope of rights and obligations of a legal entity;
- empowering AI legal personality identical to that of an individual [25].

The inconsistency of the latter assumption is obvious, since the specificity of the main essential properties of AI, (namely, methods for achieving the goals of punishment), is not taken into account. The previous assumption is also controversial, since a legal entity is an organization that owns separate property and meets its obligations, is able to have rights and bear obligations, be a plaintiff and a defendant in court. Accordingly, it is a kind of legal fiction, objectively manifested as a record in the register and acts as a collective organization established for the purpose of civil relations. Such an interpretation of AI may be valuable from the standpoint of the need to form a list of penalties that can be applied to AI, in terms of the concept of legal personality, it seems inconsistent.

As previously noted, the key feature of AI – the ability to self-learning, resulting in strong-willed and intellectual autonomy. Accordingly, consequently, it is inappropriate to raise the question of the possibility of an individual having any kind of relation to AI, to anticipate the cause and effect of its activities, as well as to predict its behavior or specific decisions. AI is a separate entity, different from a person, and even more so is not similar to the fiction created for organizing the activities of a group of people [26, 27].
Accordingly resolving the issue of his criminal responsibility is necessary to distance itself from existing concepts, since most of them have an anthropocentric orientation [31-37]. You can agree with the point of view of E.N. Barkhatova [1] that criminal responsibility should be understood have emerged since the indictment and continuing until cessation of all adverse criminally-legal consequences of the duty of a person in criminal legal relations be responsible for act committed by him, containing signs of a crime, and to endure the related measures of criminal procedural coercion measures under criminal law, the negative consequences of the conviction, to be punished in order to restore social justice, crime prevention and corrections of the face [26-29].

It would be appropriate to enshrine the proposed definition of criminal liability in the criminal law in order to bring about uniformity in its interpretation and application, and thus to respect the principle of justice. Criminal liability is implemented in forms that can be divided into three groups: punishment, trials and warnings. In its theoretical essence, criminal responsibility is a type of legal responsibility, which is a statutory obligation to bear responsibility for crimes committed and is expressed in the application of the authorized state body (court) of criminal punishment and other measures of a criminal legal nature [9]. Accordingly, the forms of its implementation can be: the appointment of criminal punishment; application of compulsory measures of educational influence; application of compulsory measures of medical character. In other words, criminal liability entails a negative change in the legal status 20, a reduction in the scope of the rights of the subject in order to restore social justice.

It seems to be an indisputable judgment that the identity of the criminal responsibility of man and AI is impossible. However, abstaining from its establishment is also impractical. Affected in the context of a possible criticism of our position, based on the fact that one of the constants of criminal responsibility is the awareness by a person undergoing criminal liability, the prohibition of their behavior, the existence of a causal link between his criminal actions negative changes in the protected criminal law to public relation and the application to it of state interventions, resulting in the formation of that person's respect for law and society, the elimination of persistent antisocial installation, prevention of new crimes in the future, which implies sanity and consciousness. Accordingly, there are reasonable grounds to believe that AI is capable of conscious behavior [25].

In support of our position, we give the following argument: one of the determinants of awareness is the ability to reflect adequately the surrounding reality and intellectual development, which consists in the ability to self-study. Let's confirm the presence of AI capabilities with the following example.

On December 7, 2017, Google's Alpha Zero program defeated AI Stock fish, the champion of computer programs. While mastering chess, Alpha Zero used the most advanced machine learning techniques, playing with itself. However, out of the hundred games that were played with Stock fish Challenger, Alpha Zero won 28 and squared a game for 72. At the same time, we focus on the fact that the program did not contact a person during training, and the duration of the program training since the end of programming was 4 hours [30]. Thus, statements about the lack of AI consciousness or the ability to self-learning seem to be untenable.

In addition, in this situation, another example is possible, which clearly demonstrates the ability of an information algorithm (not AI) to be exposed to third-party influence and change its behavior. In the Instagram app, in the panel at the bottom of the device screen (version 113.0) [3], the second command on the left (indicated by a magnifying glass icon) allows you to go to the suggested section, after which a group of publications is observed on the device screen. You can view this group by moving your index finger up from the bottom of the screen. If the "Select publication for subject viewing" command is given (by clicking on the corresponding area of the screen), the publication is displayed in a magnified form.
4 Results

At the same time, in the upper right corner there is a label in the form of three parallel dots, when you navigate through it, a dialog box with the following commands is displayed: "Complain/see fewer of such publications". When you click on the area of the screen that indicates the command "See fewer of such publications", a dialog box with a textual inscription of the following contents is displayed: "Publication is hidden. Now you will see fewer such publications" [13]. When applying this algorithm repeatedly to publications of the same kind, it is easy to notice that they will disappear from the "Suggested" section. Based on the above, we can conclude that even the simplest electronic algorithm that provides a selection of publications, can be subjected to measures of influence and, accordingly, to change undesirable behavior, which is similar in principle to the mechanism of criminal responsibility [11].

It is thought that, accumulating data coming from the screen of the device used, the electronic algorithm, on the basis of identifying patterns, decides on the formation of the section "Suggested". It is unlikely to be true that the physical entity has a direct impact on what publications the Instagram program will provide to a particular user. Accordingly, this decision is made autonomously based on the collection of statistical data about the user and his preferences [24].

The essence of the command "See fewer such publications" is the requirement to change the order of the algorithm in terms of forming an opinion about the user's preference, and according to which the correction of the section "Suggested" is carried out. That is, the user, submitting by clicking on the screen of the device command, changes the laws of the program, while achieving a positive result in the form of preventing the subsequent submission of unwanted publications. It seems that there are sufficient grounds for an analogy with AI and measures of influence in the form of an instruction submitted to it with the mandatory requirement to refrain from undesirable deviations in the future. Based on the above, we have concluded that AI is fundamentally capable of being criminally liable, while also effectively achieving its goals and preventing negative deviations in the future, that is, correcting its own behavior under the influence of coercive measures.

5 Discussions

The position advocated in this paper has been repeatedly expressed and consistently justified in the course of scientific discussions held in the framework of the III all-Russian youth scientific and practical conference "Investigative activities: problems, their solution, prospects for development" held by the Moscow Academy of the Investigative Committee of Russia; II all-Russian scientific and practical conference "Vector of development of management approaches in the digital economy" organized by the Kazan innovation University, as well as during the speech at the plenary session of the conference Kazan scientific readings 2019, also held at the Kazan innovation University. The opinions expressed naturally caused an active scientific discussion during which we confirmed the following provisions:

1. Artificial intelligence - is a cybernetic education, different from computer programs, capable of Autonomous, conscious-volitional behavior,

2. Artificial intelligence contains significant criminological risks in its implementation and partially controlled development, which form the need to prepare legal models for preventing criminally dangerous behavior and preparing prerequisites for legal responsibility for the actions of artificial intelligence.
3. Artificial intelligence is capable of committing acts that have a full set of features selected by the national legislator to refer it to the General concept of a crime;

4. Models of criminal responsibility of modern criminal law do not meet the requirements of developing technologies, which can lead to widespread objective imputation of the actions of artificial intelligence to developers. The vast majority of participants in scientific discussions expressed solidarity with our position at the end of the debate.

6 Finding

Having made the above research we find grounds to assert the following facts:

1) The national legal system of Russia demonstrates absolute unfitness for regulation and unfitness for regulating relations involving artificial intelligence, which is determined by the traditional understanding of the subject of law as only and exclusively human;

2) High crime risks and lack of tactics and methods of detection, prevention and suppression of the activities of the artificial intelligence forms the priority development of criminal law protection of appropriate areas of public relations, and orientation of heads of law enforcement agencies for training tactical recommendations, methods and means to implement operational activities to neutralize threats posed by artificial intelligence;

3) Recognition of artificial intelligence as a subject of crime is a scientifically and practically justified measure caused by the objective need to ensure the security of the individual, society and the state in modern conditions of accelerated development of technologies and global digitalization of society.

7 Conclusions

In conclusion, we suggest that the research presented in this paper, the facts stated and the conclusions formulated will receive an objective assessment of the scientific community and, in the future, form the basis of the initial theoretical understanding of the prerequisites for legal regulation of relations with artificial intelligence on the basis of a deviation from the generally recognized, fundamental, but not meeting modern technological challenges, principle of recognizing the subject of relations only and exclusively human. In turn, discussions about potential threats will not be ignored and will stimulate scientific discussion and regulatory regulation. We hope that the results of this work will allow many to re-evaluate their own position on artificial intelligence and agree that it has properties and qualities that exclude the extension of the legal regime of the object of law to it.

References


8. The newest drone "hunter" made the first joint flight with the su-57. RIA "Novosti". Retrieved from: https://ria.ru/20190927/1559199891.html


