

Conceptual Bases of Criminal Punishment Application Regarding Artificial Intelligence

Bases conceptuales de la aplicación de la sanción penal en relación con la inteligencia artificial

Authors

Ramil Rustamovich Gaifutdinov, Ildar Rustamovich Begishev, Zarina Ilduzovna Khisamova, Elina Leonidovna Sidorenko, Diana Davletovna Bersei, Regina Rustemovna Musina

¹⁾ Department of Criminal Law, Kazan Federal University, Kazan, Russia (Russian Federation).kafedra.ksu@yandex.ru

²⁾ Kazan Innovative University named after V.G. Timiryasov, Kazan, Russia (Russian Federation).begishev@mail.ru

³⁾ Department of Planning and Coordination of Research Activities, Krasnodar University of the Ministry of Internal Affairs of the Russian Federation, Krasnodar, Russia (Russian Federation).alise89@inbox.ru.

⁴⁾ Department of Criminal Law, Criminal Procedure and Criminalistics, Moscow State Institute of International Relations (University), Moscow, Russia (Russian Federation).12011979@list.ru

⁵⁾ Faculty of Law, Department of Criminal Law and Procedure, North-Caucasus Federal University .Stavropol, Russia (Russian Federation).di.bersej2012@yandex.ru

⁶⁾ Department of Criminal Law and Procedure, Kazan Innovative University named after V.G. Timiryasov, Kazan, Russia (Russian Federation).unabhangig@yandex.ru

Fecha de recibido: 2020-11-30

Fecha de aceptado para publicación: 2021-02-01

Fecha de publicación: 2021-03-25





Abstract

At present, the widespread development and use of artificial intelligence technologies (hereinafter referred to as AI) set tasks for the world community, the examples of the solution of which were not previously available in the history of mankind. One of the priority areas of the legislator activity in the foregoing field is the creation of the foundations, principles and mechanism (concept) of criminal penalty application in relation to AI. At the moment, there are no legal orders of any kind in this area, the topic of punishment application to AI is also not widely used in criminal law literature, which seems somewhat illogical in terms of legal regulation timeliness, completeness and effectiveness. It seems that the currently existing conceptual framework for the application of criminal penalties in relation to AI is not able to have an effective impact on AI, due to its specific material manifestation, which we tend to subdivide into three types. The first of them is the AI unit - a material object of the world, which is an integral part of the AI and designed to implement its functionality. The next in order is AI, embedded in a material carrier, which is an object of the material world itself and is not able to carry out functional actions aimed at AI potential implementation. And the last of the named varieties are AI, which has an expression in the cybernetic world that is not directly objectified into the surrounding reality.

Keywords: artificial intelligence, types of artificial intelligence, criminal punishment, punishments, types of artificial intelligence punishments, criminal liability, criminal law.

Resumen

En la actualidad, el desarrollo y uso generalizados de tecnologías de inteligencia artificial (en adelante, IA) planteó tareas a la comunidad mundial, cuyos ejemplos de solución no estaban disponibles anteriormente en la historia de la humanidad. Una de las áreas prioritarias de la actividad del legislador en el ámbito anterior es la creación de los fundamentos, principios y mecanismo (concepto) de aplicación de la pena penal en relación con la IA. Por el momento, no existen órdenes legales de ningún tipo en esta área, el tema de la aplicación del castigo a la IA tampoco es muy utilizado en la literatura de derecho penal, lo que parece algo ilógico en términos de oportunidad, integridad y efectividad de la regulación legal. Parece que el marco conceptual existente en la actualidad para la aplicación de sanciones penales en relación con la IA no es capaz de tener un impacto efectivo sobre la IA, debido a su manifestación material específica, que tendemos a subdividir en tres tipos. El primero de ellos es la unidad de IA, un objeto material del mundo, que es una parte integral de la IA y está diseñado para implementar su funcionalidad. El siguiente en orden es la IA, incrustada en un portador de material, que es un objeto del mundo material en sí mismo y no puede llevar a cabo acciones funcionales destinadas a la implementación potencial de la IA. Y la última de las variedades nombradas es la IA, que tiene una expresión en el mundo cibernético que no se objetiva directamente en la realidad circundante.

Palabras clave: inteligencia artificial, tipos de inteligencia artificial, sanción penal, penas, tipos de sanciones de inteligencia artificial, responsabilidad penal, derecho penal.



Introduction

Artificial intelligence (AI) is a technological term that refers to objects used in response to detected contexts to detect contexts or to affect behaviour. Our ability to create such objects has evolved, and with it, the effect they have on our society. AI technology is already being implemented in various social practices, for example, in the field of transport infrastructure facility use designed for vehicle movement, which is expressed in the creation of unmanned vehicles and entails the formation of some legal uncertainty regarding their place in the structure of legal relations (Hildebrandt, 2018).

In addition to the abovementioned, the special significance of AI is indicated by the Decree of the RF President (May 7, 2018) "On National Goals and Strategic Development Tasks for the Period until 2024" (Epinina et al., 2019; Pleslov & Lukashenko, 2020) and "Strategies for the development of the information society in the Russian Federation for 2017-2030" (Inshakova et al., 2019; Kail et al., 2019; Kulikova et al., 2019; Tarakanov et al., 2019), where the development of a system of the digital economy and AI legal regulation was defined as a priority task of the digital economy development (Kutyłowski et al., 2020; Vasiliev et al., 2019). Concerning this vector of legal regulation development, reasonable opinions are expressed by science that it is at the stage of origin in Russia (Sitdikova et al., 2015).

1.1 Objective

Formulating the philosophical framework for the implementation of criminal penalties in relation to AI, we start from the assumption that there should be no shift in the aims and core principles of punishment because they have a pattern that is implemented rather than driven.

Material And Methods

Artificial intelligence is being incorporated seamlessly into our daily lives, enhancing our driving awareness and skills, avoiding traffic, meeting friends, choosing the ideal movie, and even cooking a healthy meal. It also has a major impact on many areas of society and industry, ranging from technological exploration to smart cities, transportation and development to healthcare and medical diagnostics. The materials for the work were the articles posted in scientific journals and on websites, as well as Russian and foreign legal acts.

The methodological basis of the study is the combination of methods of scientific knowledge, including abstract logical method, comparison and correlation analysis.

Results And Discussion

Along with this, AI introduction into public relations is necessary to assess the possible risks that these changes may involve, the use of AI will entail the need for a radical review of existing anthropocentric concepts of legal regulation, which is recognized by scientists as an objective necessity, otherwise the proper level of effective legislative impact on social practices with the participation of AI is not possible (Z. I. Khisamova & Begishev, 2019a), due to the fact that the activities for its implementation and use are sufficiently specific so that its essential properties, which can create criminological risks, not put into a safe state by the existing legal system (Ildar R. Begishev & Khisamova, 2018). At that, one should also mention the correct remark by O.S. Kapinus, who indicates that in modern conditions of digital technology development, along with the timely detection of technology danger for society and the state, it should also be soberly assessed and potentially progressive innovations should not be blocked by legal prohibitions (Kapinus, 2018).

One of the priority areas of the legislator activity in the foregoing field is the creation of the foundations, principles and mechanism (concept) of criminal penalty application in relation to AI. At the moment, there are no legal orders of any kind in this area, the topic of punishment application to AI is also not widely used in criminal law literature, which seems somewhat illogical in terms of legal regulation timeliness, completeness and effectiveness (Bikeev et al., 2019).

In addition to the foregoing, the scientific community also questions the possibility of AI to aware of causal relationship development between an act and the consequences that it entails, as well as the likelihood of their occurrence (Buyers & Partner, 2015). Allegations are made that the actions of AI must entail criminal responsibility of its creators or users (Nevejans, 2016). The model of the arguments by M.T. Jones is very informative in the affected context, in terms of AI behaviour calculation in the field of computer simulation (Jones, 2004).

Developing our reasoning, it is necessary to pay special attention to the relevance of bringing a more comprehensive description of AI from the perspective of digital technologies (Efremova, 2020), including bringing a reasonable classification that will be necessary for further reasoning.

So, depending on the intellectual potential, it is customary to distinguish between a strong AI, which is software, thanks to which computers can think as people do, and also receive other abilities of a rational being, and a weak AI that can carry out a single type of activity, a strictly defined list of functions for which it was programmed (for example, to provide a user with relevant and acceptable information data, as will be indicated below). Along with the abovementioned, it is necessary to identify the broader classification set



forth in the report of Cybercrime Observatory at the Australian National University (Broadhurst et al., 2019), according to which AI is divided into three categories (Weak AI / Medium AI / Strong AI). However, there are no fundamental differences in the differentiation given, since the criterion for distinguishing between Weak AI and Medium AI is the completeness and effectiveness of contact during interaction with a person; Strong AI, in turn, is distinguished by its self-learning ability. It is necessary to focus attention on this ability since the possibility of self-learning involves independent, autonomous identification of new knowledge, based on the ability to perceive, process, accumulate and use information from the outside world.

In the affected context, the position of scientists seems to be quite informative and relevant to the subject of cognition when they indicate that AI is a mathematical model per se located in a certain technical device, which is a self-learning model of neural connections (Doroganov & Baumgarten, 2013) capable of perceiving objective reality information, processing it, and thus gain new knowledge not invested by developers; accordingly, the potential for its behaviour prediction decreases significantly, or rather, it moves from the plane of control over the created object to the area of an autonomous subject behaviour prediction, and therefore, there is a certain section in modern computer science that develops artificial neural networks for their subsequent use in AI systems. In addition to the abovementioned, scientists make an important observation, according to which artificial neural networks should be understood as mathematical models capable of learning, and created like the human brain.

The aforesaid means that creation of AI with limited functionality (a list of knowledge and possible behaviours) in the future makes it possible to detect its significant, progressive development. In this case, it is hardly advisable to pose the question of AI behaviour prediction certainty by a person who created it. We do not deny the possibility of AI behaviour prediction. However, we note that its certainty will be comparable to the prediction of an autonomous subject behaviour of independent will, and not an instrument or a tool that is in the physical possession of a subject and can cause harm only indirectly, by committing certain actions with it. During the determination of the causal relationship between AI actions and socially dangerous consequences caused in each case, it is necessary to exclude the possibility of unauthorized access to AI for third parties (Z. I. Khisamova & Begishev, 2019a). A significant number of works have been devoted to the study of these issues, as well as the aspects of the criminal law and civil law regulation in respect of public relations arising from the use of AI (I. R. Begishev, 2020; I. R. Begishev et al., 2019; Ildar R. Begishev et al., 2019, 2020; Bokovnya et al.,

2019; Hallevy, 2015; Z. Khisamova et al., 2019; Z. I. Khisamova & Begishev, 2019b; Zarina I. Khisamova et al., 2019; Latypova et al., 2019; Mosechkin, 2019; Shestak et al., 2019; Shestak & Volevodz, 2019; Simmler & Markwalder, 2019; Sukhodolov et al., 2020; Sukhodolov & Bychkova, 2018; Uzhov, 2017).

It seems that the currently existing conceptual framework for the application of criminal penalties in relation to AI is not able to have an effective impact on AI, due to its specific material manifestation, which we tend to subdivide into three types. The first of them - the AI unit - a material object of the world, which is an integral part of the AI, and designed to implement its functionality. The next in order is AI, embedded in a material carrier, which is an object of the material world and is not able to carry out functional actions aimed at AI potential implementation. And the last of the named varieties are AI, which has an expression in the cybernetic world that is not directly objectified into the surrounding reality.

We named the classification to provide further discussion of the proposed types of punishment since the mechanism of their application directly correlates with the material expression of AI in the world around us. It is unlikely that the physical destruction of AI, which is a cybernetic entity, will be deemed appropriate.

Conclusion

Formulating the conceptual foundations for the application of criminal penalties in relation to AI, we proceed from the belief that the goals and main principles of sentencing should not undergo changes, since they have an applied rather than orientating trend. Moreover, we believe that it is possible in subsequent discussions to rely on them when substantiating our own position. So, according to the provisions of the Art. 60 of the RF Criminal Code, it is necessary to take into account a sentence influence on the correction of a guilty party when sentencing - AI in the case under consideration. In the affected aspect, we may be objected that the ability of AI to correction is not reliably established and can be criticized and questioned.

When they concentrate on the effectiveness of criminal penalty application in relation to AI, for the purpose of its correction, one should take into account the stability of the antisocial attitude of the aforementioned cyber education, which may be caused, in particular, by the process of its creation, in other words, the conditions and circumstances of a particular AI creation should be taken into account, the goals of its programming should be established in each case, whether it was purposefully created to commit crimes.



In the process of AI use, the following situations are possible that require criminal law regulation. Firstly, an AI with the ability to self-study can autonomously decide to commit an act qualifying as a crime, and secondly, an AI can be created to commit crimes. What we see is a fundamental difference, it is unlikely that application of the penalty to the AI programmed originally for committing crimes not related to its destruction will fully protect society from the commission of similar crimes in the future by the indicated cybernetic formation. It seems that truncating the properties of AI, and applying a less severe punishment on the basis and within limits provided by the criminal law, will not be able to provide sufficient guarantees of social stability. It should further be noted that there may be some difficulty in determining the justice of the punishment applied to the AI, that is, its compliance with the committed crime nature and degree of public danger, the circumstances mitigating and aggravating the punishment, the totality of AI essential properties, which determine the tendency of its behaviour in terms of moral assessment (a similar wording is given by us as a substitute for the concept of "the perpetrator identity"). In order to overcome this uncertainty, let's note that, as was mentioned above, due to the presence of intellectual and volitional elements in AI actions, as well as the ability to perceive correctly the circumstances that have legal significance and understand them (this follows from the presence of neural networks in AI simulating the human brain, respectively, it does not exclude its ability to consciously volitional behaviour, and to the abovementioned possibilities). There is no fundamental difference in the assessment of the indicated characteristics of sentencing to a person and to AI.

Acknowledgements

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

References

- Begishev, I. R. (2020). Organization of the hacker community: Criminological and criminal law aspects. *All-Russian Criminological Journal*, 14(1), 96–105.
- Begishev, I. R., Khisamova, Z. I., & Mazitova, G. I. (2019). Information infrastructure of safe computer attack. *Helix*, 9(5), 5639.
- Begishev, Ildar R., & Khisamova, Z. I. (2018). Criminological Risks of Using Artificial Intelligence. *RUSSIAN JOURNAL OF CRIMINOLOGY*, 12(6), 767–775.
- Begishev, Ildar R., Khisamova, Z. I., & Mazitova, G. I. (2019). Criminal legal ensuring of security of critical information infrastructure of the Russian Federation. *Revista Gênero & Direito*, 8(6), 283–292.
- Begishev, Ildar R., Latypova, E. Y., & Kirpichnikov, D. V. (2020). Artificial Intelligence as a Legal Category: Doctrinal Approach to Formulating a Definition. *Actual Probs. Econ. & L.*, 79.
- Bikeev, I., Kabanov, P., Begishev, I., & Khisamova, Z. (2019). Criminological risks and legal aspects of artificial intelligence implementation. *Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing*, 1–7.
- Bokovnya, A. Y., Khisamova, Z. I., & Begishev, I. R. (2019). Study of Russian and the UK Legislations in Combating Digital Crimes. *Helix*, 9(5), 5458–5461.
- Broadhurst, R., Maxim, D., Brown, P., Trivedi, H., & Wang, J. (2019). Artificial Intelligence and Crime. *Available at SSRN 3407779*.
- Buyers, J., & Partner, O. C. L. (2015). Liability Issues in Autonomous and Semi-Autonomous systems. *Osborne Clarke LLP*.
- Doroganov, V. S., & Baumgarten, M. I. (2013). Possible problems arising during artificial intelligence development. *Vestnik Kuzbasskogo Gosudarstvennogo Tekhnicheskogo Universiteta*, 4, 98.
- Efremova, M. A. (2020). Crimes in the Sphere of Digital Information Circulation. *Actual Probs. Econ. & L.*, 201.
- Epinina, V. S., Popova, A. R., & Popova, O. Y. (2019). GR Management Actualization in the Context of Achieving the National Development Goals of the Russian Federation for the Long Term. *Competitive Russia: Foresight Model of Economic and Legal Development in the Digital Age. International Scientific Conference in Memory of Oleg Inshakov*, 267–275.
- Hallevy, G. (2015). *Liability for crimes involving artificial intelligence systems*. Springer.
- Hildebrandt, M. (2018). Law as computation in the era of artificial legal intelligence: Speaking law to the power of statistics. *University of Toronto Law Journal*, 68(supplement 1), 12–35.
- Inshakova, E. I., Ryzhenkov, A. Y., & Inshakova, A. O. (2019). Neo-industrialization of the Russian economy: Technological and digital development. In *Ubiquitous Computing and the Internet of Things: Prerequisites for the Development of ICT* (pp. 239–250). Springer.
- Jones, M. T. (2004). *Artificial intelligence programming in applications*. M.: DMK Press.



- Kail, Y. Y., Shokhnekh, A. V., Samsonova, M. V., Lamzin, R. M., & Elsukova, Y. Y. (2019). Prospects of GR Management Development in the context of forming a digital economy. *Prospects*, 40(22), 30.
- Kapinus, O. (2018). Criminalization and decriminalization of acts: Finding the best balance. *Obshchestvennye Nauki i Sovremennost*, 4, 37–46.
- Khisamova, Z., Begishev, I., & Gaifutdinov, R. (2019). On methods to legal regulation of artificial intelligence in the world. *SCOPUS-2019-9-1-SID85075304864*.
- Khisamova, Z. I., & Begishev, I. R. (2019a). Criminal liability and artificial intelligence: Theoretical and applied aspects. *All-Russian Journal of Criminology*, 13(4), 574.
- Khisamova, Z. I., & Begishev, I. R. (2019b). Legal regulation of artificial intelligence. *Baikal Research Journal*, 10(2).
- Khisamova, Zarina I., Begishev, I. R., & Sidorenko, E. L. (2019). Artificial Intelligence and Problems of Ensuring Cyber Security. *International Journal of Cyber Criminology*, 13(2), 564–577.
- Kulikova, N. V., Persteneva, N. P., & Ruslanova, T. V. (2019). Information Society Development in Regions of the Russian Federation. *International Scientific Conference "Digital Transformation of the Economy: Challenges, Trends, New Opportunities"*, 214–224.
- Kutyłowski, M., Lauks-Dutka, A., & Yung, M. (2020). GDPR—Challenges for Reconciling Legal Rules with Technical Reality. *European Symposium on Research in Computer Security*, 736–755.
- Latypova, E. Y., Nechaeva, E. V., Gilmanov, E. M., & Aleksandrova, N. V. (2019). Infringements on Digital Information: Modern State of the Problem. *SHS Web of Conferences*, 62, 10004.
- Mosechkin, I. N. (2019). Artificial Intelligence and Criminal Liability: Problems of Becoming a New Type of Crime Subject. *Vestnik Saint Petersburg UL*, 461.
- Nevejans, N. (2016). *European Civil Law Rules in Robotics: Study*. Publications Office.
- Pleslov, A., & Lukashenko, O. (2020). Regional Aspects Of 2024 National Projects on Strategic Development in Russia. *Economic and Social Development: Book of Proceedings*, 77–86.
- Shestak, V. A., & Volevodz, A. G. (2019). Modern requirements of the legal support of artificial intelligence: A view from Russia. *Russian Journal of Criminology*, 13(2), 197–206.
- Shestak, V. A., Volevodz, A. G., & Alizade, V. A. (2019). On the possibility of doctrinal perception of artificial intelligence as the subject of crime in the system of common law: Using the example of the US criminal legislation. *Russian Journal of Criminology*, 13(4), 547–554.
- Simmler, M., & Markwalder, N. (2019). Guilty Robots?—Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence. *Criminal Law Forum*, 30(1), 1–31.
- Sitdikova, L. B., Shilovskaya, A. L., Volkova, M. A., Lenkovskaya, R. R., & Stepanova, N. A. (2015). Legal regulation and copyright protection in Internet in Russia and abroad. *Mediterranean Journal of Social Sciences*, 6(6), 163.
- Sukhodolov, A. P., Bychkov, A. V., & Bychkova, A. M. (2020). *Criminal Policy for Crimes Committed Using Artificial Intelligence Technologies: State, Problems, Prospects*.
- Sukhodolov, A. P., & Bychkova, A. M. (2018). Artificial intelligence in crime counteraction, prediction, prevention and evolution. *Russian Journal Of Criminology*, 12(6), 753–766.
- Tarakanov, V. V., Inshakova, A. O., & Dolinskaya, V. V. (2019). Information society, digital economy and law. In *Ubiquitous Computing and the Internet of Things: Prerequisites for the Development of ICT* (pp. 3–15). Springer.
- Uzhov, F. V. (2017). Artificial intelligence as a subject of law. *Probely v Rossiiskom Zakonodatel'stve*, 3, 357–360.
- Vasiliev, A., Muratkhanova, M., & Voronkova, E. (2019). Law in the digital age: Regulating a new technological reality. *International Conference on Sustainable Development of Cross-Border Regions: Economic, Social and Security Challenges (ICSDCBR 2019)*.