

ТЕОРЕТИКО-ПРАВОВЫЕ АСПЕКТЫ СИСТЕМЫ НАКАЗАНИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

И.Н. Романова

ЧОУ ВО «Московский университет имени С.Ю. Витте», г. Рязань

Целью настоящей статьи является исследование вопросов необходимости теоретико-правовой разработки системы наказаний искусственного интеллекта ввиду того, что неконтролируемые кибернетические творения несут в себе потенциальные и вполне вероятные риски наступления нежелательных последствий. В рамках работы решаются задачи по анализу социальной практики применения технологий искусственного интеллекта; оценке рисков и возможных угроз использования интеллектуальных информационных систем; внесению предложений по недопущению возможных негативных последствий практической реализации интеллектуальных цифровых ресурсов, а также нивелированию наступивших нежелательных последствий; формулированию и обоснованию концепции наказаний за действия искусственного интеллекта. Методологическую основу исследования составили общенаучные методы познания, а также юридико-технический частнонаучный метод. Результатом проведенного исследования является формулирование и обоснование позиции, согласно которой применение системы наказаний искусственного интеллекта в любом из его материальных проявлений, способной выступать в качестве меры принудительного характера, представляется целесообразным. Выводом автора по итогам рассмотрения избранной проблематики является утверждение, что преодоление в механизме практической реализации системы наказаний искусственного интеллекта вероятных препятствий, не носящих принципиально непреодолимого характера, напрямую зависит от приложенных усилий.

Ключевые слова: вероятные криминологические риски, программа искусственного интеллекта, развитие цифровой экономики, уголовно-правовая ответственность, угроза общественной безопасности.

В условиях современных реалий сложным представляется бесперебойное функционирование целого ряда процессов в самых различных отраслях экономики и сферах жизни общества вне применения цифровых технологий, способных обеспечивать позитивные трансформации в заданных направлениях. Крупные корпорации (в частности, транснациональные компании) в настоящее время заинтересованы в технологиях искусственного интеллекта, которые способны оказать существенное влияние на многие сферы человеческой деятельности, включая экономику и политику.

На мировом рынке искусственного интеллекта существуют следующие группы субъектов:

- компании, которые разрабатывают технологии искусственного интеллекта и имеют собственные данные (например, Google, Microsoft);
- компании, деятельность которых связана с обработкой больших объемов данных (например, IBM, Oracle);
- инновационные компании, которые решают конкретные проблемы, но не имеют собственных данных.

Однако справедливым будет отметить, что использование интеллектуальных информационных систем имеет двоякий эффект: с одной стороны, оно дает значительные экономические выгоды тем, кто его использует, а с другой стороны, оно способно внести значительные негативные аспекты для общества в целом. Таким образом, совершенно очевидными становятся определенные риски наступления вероятных неблагоприятных и нежелательных последствий в результате неконтролируемого применения технологий искусственного интеллекта, способных выступить прямой или косвенной угрозой для устойчивого развития общества и безопасности государства, поскольку, как и любое открытие, искусственный интеллект может быть использован для различных целей. Например, динамит был создан для добычи полезных ископаемых, и никто не думал, что он может найти негуманное применение.

В социальной практике известны случаи причинения вреда искусственным интеллектом. Так, например, в марте 2018 г. автомобиль компании Uber, пилотируемый искусственным интеллектом, в результате сбоя оборудования сбил в Сан-Франциско женщину, двигавшуюся перпендикулярно проезжей части, что повлекло за собой ее смерть. Было установлено, что интеллектуальная программа управления автомобилем ошибочно расценила это как «ложную операцию». Другим, не менее показательным примером, является эпизод, произошедший в 2016 г. в сети Интернет, когда чатбот Тэй, созданный компанией Microsoft при поддержке китайского поисковика Baidu, запрограммированный повторять слова людей, не справился с натиском интернет-пользователей, провоцировавших программу на поддержку с ними диалога в ключе дискриминирующих и оскорбляющих высказываний (встроенный в программу фильтр вульгарных выражений вышел из строя), и менее чем через 24 часа после запуска данной программы начал поддерживать Гитлера и формулировать расистские и шовинистские высказывания [8].

В целях, как минимум, снижения и, как максимум, нивелирования риска ущерба, причинение которого возможно вследствие применения технологий искусственного интеллекта, их внедрение и практическая реализация не только возможны, но и должны быть осуществлены

исключительно в рамках соблюдения соответствующих необходимых мер предосторожности.

В целом, программу искусственного интеллекта можно представить как не имеющее материального выражения творение кибернетики, никак не выражающееся в материальном мире, существующее исключительно в цифровом пространстве, но по своим качествам способное совершать действия, влекущие за собой уголовно-правовые последствия с применением мер неблагоприятного воздействия.

Как отмечается в научных источниках, наказание реализует общую и частную уголовно-правовую профилактику и защищает общество от субъектов, которые представляют угрозу причинения вреда общественным отношениям, охраняемым уголовным законом [3].

Наказание содержит два элемента: юридический – лишение или ограничение прав и свобод, которое заключается в лишении человека возможности реализовать определенный тип и меру возможного поведения (объем действий), вытекающий из отчужденного права или свободы, который используется для восстановления социальной справедливости, исправления осужденного и предотвращения совершения новых преступлений; и фактический – создание условий, и фактических обстоятельств, при которых осуществление этого права будет невозможно. Основопологающим является следующее: исходя из позиции суда, наказание – это лишение или ограничение права или свободы, из которого вытекает определенный тип и мера возможного поведения человека, гарантированная законом, но отчужденная от него как мера государственного принуждения, которая выражается в создании условий, исключающих возможность осуществления права, т. е. возможность действовать определенным образом. Вышесказанное явно подразумевает возможность применения наказания искусственного интеллекта, поскольку единственной трудностью может быть создание условий (т. е. фактического элемента), а не принципиальная невозможность применения уголовного наказания к искусственному интеллекту как правовая модель, которая, напротив, кажется разумной и вполне возможной с юридической точки зрения. Соответственно, применение наказаний искусственного интеллекта в любом из его материальных проявлений представляется мотивированным и целесообразным, и они носят исключительно практический характер, а их преодоление зависит от приложенных усилий и не имеет принципиально непреодолимых препятствий [1].

Вместе с тем полное отождествление механизма наказания, предусмотренного для человека, и наказания для искусственного интеллекта представлялось бы неправильным, поскольку в случае применения репрессивных мер воздействия на человека генеральной целью является исправление виновного, в отношении которого вступил обвинительный приговор суда, во втором же случае – при

ответственности искусственного интеллекта – ставится цель по обеспечению общественной безопасности, безопасности как отдельно взятой личности, так и в целом государства от деятельности искусственного интеллекта, носящей противоправный характер, а также цель по предотвращению совершения новых преступлений. Не стоит категорически исключать версию о том, что отдельные виды неблагоприятного наказательного воздействия, применяемые к искусственному интеллекту, способны сформировать у него убеждение в неприемлемости и недопустимости поведения такого рода. В то же время отсутствуют и достаточные основания, позволяющие определенно полагаться на желаемый положительный результат. Также стоит отметить, что формулирование видов наказаний в отношении искусственного интеллекта не обременено ограничительными рамками принципов и норм гуманистического характера, являющихся непреложными для людей и признанными на международном уровне [2].

Концепция наказания искусственного интеллекта представляется сложно совместимой с универсальными концепциями уголовного права. Тем не менее наказание искусственного интеллекта нельзя категорически исключать с помощью быстрых теоретических аргументов, основанных на аналогиях с корпоративной и строгой уголовной ответственностью, а также знакомых понятиях вменяемости и вменения. Наказание искусственного интеллекта может привести к общему сдерживанию и выразительным преимуществам, а негативные ограничения, такие как наказание за превышение вины, не должны нарушаться. К числу видов наказания искусственного интеллекта можно отнести следующие: полная деактивация (уничтожение); частичная деактивация; конфискация в пользу государства; запрет на применение в определенных сферах деятельности; осуществление безвозмездной деятельности в интересах государства или общества; придание статуса осужденного (неблагонадежного) [5].

Таким образом, в условиях современных реалий, которые уже не представляются возможными вне развития цифровой экономики, мировому сообществу в целях обеспечения эффективной защиты от неправомерных воздействий со стороны искусственного интеллекта следует осуществить разработку системы наказаний за действия искусственного интеллекта [4].

В исследовательских целях и основываясь на изложенном, систему наказаний искусственного интеллекта возможно определять следующим образом.

Полная деактивация (уничтожение) – это принудительное воздействие на нейронные связи искусственного интеллекта. Таким образом, они полностью теряют свои свойства. Данный вид наказания можно рассматривать как наиболее репрессивный. Представляется, что полная деактивация (уничтожение) искусственного интеллекта будет

являться справедливой мерой неблагоприятного воздействия в случае совершения искусственным интеллектом тяжкого либо особо тяжкого преступления [6].

Следующим видом наказания с точки зрения репрессивного потенциала является частичная деактивация искусственного интеллекта – перевод в категорию более слабого искусственного интеллекта путем его потенциального интеллектуального снижения. Указанный вид наказания, на наш взгляд, целесообразно осуществлять путем принудительного воздействия на нейронные связи искусственного интеллекта. Таким образом, способность обрабатывать определенное количество информации из внешнего мира может быть устранена. В то же время часть информации, необходимой для реализации социально значимых функций, будет доступна, но явно недостаточна для сознательно-волевого автономного поведения. Так, при решении вопроса о применении данного вида наказания суды должны установить экономическую целесообразность продолжения функционирования искусственного интеллекта. Представляется возможным сопоставить указанный вид наказания с конкретным испытательным сроком, в течение которого деятельность искусственного интеллекта будет постоянно контролироваться надзорным органом.

Конфискация искусственного интеллекта в пользу государства – это принудительное его изъятие у владельца, который им распоряжается, и обращение его в пользу государства для выполнения полезных общественно-государственных функций. Указанный вид наказания еще в большей степени зависит от экономической целесообразности поддержания функционирования искусственного интеллекта, чем предыдущий. Однако, учитывая возможную готовность искусственного интеллекта к совершению преступлений или неполное устранение антиобщественного отношения, целесообразно использовать его в случае совершения преступления небольшой или средней тяжести в первый раз или в результате случайного стечения обстоятельств. Также следует учитывать, что на современном этапе развития каждый искусственный интеллект представляет значительную научную, промышленную, оборонную и в равной степени различную социальную или государственную ценность. Эти обстоятельства обусловили необходимость определить такой вид наказания для искусственного интеллекта, как конфискация в пользу государства. Думается, что было бы неразумно полностью уничтожить искусственный интеллект или значительно уменьшать его свойства, когда искусственный интеллект совершил преступление, не имеющее достаточной степени общественной опасности; вполне возможно добиться цели наказания способами, не связанными с причинением вреда таким кибернетическим творениям [7].

Интересным будет рассмотреть особенности запрета на использование искусственного интеллекта в определенных сферах деятельности, который может выглядеть как лишение права занимать определенные должности или заниматься определенной деятельностью в качестве наказания, применяемого к человеку. Указанное наказание целесообразно применять только в тех случаях, когда искусственный интеллект совершил правонарушение в определенных видах деятельности (например, он проводил неправильную диагностику пациентов, нарушал правила дорожного движения и т. д.). В то же время возникает необходимость исправить эти правонарушения и разработать базу данных для принятия решения о запрете функционирования в определенных областях.

Осуществление искусственным интеллектом безвозмездной деятельности в интересах государства или общества – это принудительное выполнение искусственным интеллектом общественных работ. Указанный вид наказания в большей степени направлен на материальное восстановление социальной справедливости и может быть использован в качестве дополнительного вида наказания. Другими словами, основная ценность этого вида наказания заключается в том, что в случае совершения кибернетическим организмом преступления небольшой или средней тяжести, ему предоставляется возможность компенсировать потери собственным трудом. В то же время поддерживается баланс между устранением затрат на исполнение приговора и возмещением ущерба, причиненного деятельностью искусственного интеллекта. Чтобы прояснить суть этого вида наказания, его вполне можно сравнить с обязательными работами. Разница, однако, заключается в том, что человек осуществляет свое собственное перевоспитание при выполнении последнего, и это является целью наказания. В случае с искусственным интеллектом основное внимание уделяется компенсации за причиненный материальный ущерб.

Придание искусственному интеллекту статуса осужденного (ненадежного). Указанное негативное правовое последствие не может быть отнесено к видам наказания с достаточным обоснованием. Это можно отнести к другим мерам уголовно-правового характера из-за малого репрессивного потенциала. Эта мера заключается в обязательном информировании контрагента искусственного интеллекта о факте совершения им преступления, что неизбежно повлечет за собой репутационные потери для искусственного интеллекта и будет иметь ощутимый сдерживающий фактор, особенно в сфере коммерческой деятельности, тем самым в полной мере осуществляя общую и частную профилактику. Указанная мера представляется разумной и целесообразной для применения только в случае незначительного правонарушения [3].

Резюмируя изложенное выше, представляется возможным формулирование следующих выводов. В силу того, что технологии искусственного интеллекта несут в себе потенциальные и вполне вероятные риски наступления не желаемых последствий, сопряженных с внедрением и практической реализацией таких технологий, то видится объективно существующей необходимостью создания юридических моделей предупреждения и предотвращения опасного поведения кибернетического организма, а также условий для применения мер юридической ответственности за опасные нарушения, совершаемые искусственным интеллектом. Таким образом, правовое признание искусственного интеллекта субъектом преступления является мерой, обоснованной как с научной, так и с практической точки зрения, что обусловлено объективно существующей потребностью обеспечения в условиях современной глобальной цифровизации безопасности отдельно взятой личности, общества и в целом государства.

Список литературы

1. Антонов А.А. Искусственный интеллект как источник повышенной опасности // Юрист. 2020. № 7. С. 69–74.
2. Видергольд А.И. Уголовная ответственность искусственного интеллекта // SMART LAW FOR SMART INDUSTRY: сборник научных статей / под ред. Е.В. Титовой, Т.П. Подшивалова; Южно-Уральский государственный университет юридический институт. М.: Издательство: РГ-Пресс, 202., С. 131–137.
3. Горохова С.С. Технологии на основе искусственного интеллекта: перспективы и ответственность в правовом поле // Юрист. 2021. № 6. С. 60–67.
4. Лопатина Т.М. Защита прав человека в условиях развития цифровых технологий и формирования этики электронного государства // Защита прав и свобод человека и гражданина: современное состояние и перспективы развития: Сборник материалов Всероссийской научно-практической конференции / под ред. к.ю.н. А.А. Ходусова. М.: Международный юридический институт, 2018. С. 55–61.
5. Хафизова А.Х. Юридическая ответственность за причинение вреда системами искусственного интеллекта // Сборник научных трудов IV Международной студенческой конференции. Казань, 2021. С. 334–336.
6. Broadhurst R., Brown P., Maxim D., Trivedi H., Wang J. Artificial Intelligence and Crime // Research Paper, Korean Institute of Criminology and Australian National University Cybercrime Observatory, College of Asia and the Pacific, Canberra, June 2019. Pp. 1–70.
7. Simmler M., Markwalder N. Guilty Robots? – Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence // Criminal Law Forum. 2019. № 30. Pp. 1–31.
8. Vorontsov G. Protest of wayward robots [Электронный ресурс]. URL: <https://secretmag.ru/trends/tendencies/pyat-sluchaev-kogda-iskusstvennyi-intellekt-vzbuntovalsya.htm> (дата обращения: 02.12.2022).

Об авторе:

РОМАНОВА Ирина Николаевна – кандидат юридических наук, доцент, заведующий кафедрой гражданско-правовых дисциплин, ЧОУ ВО «Московский университет имени С.Ю. Витте», Филиал (390013, г. Рязань, Первомайский просп, д.62), ORCID: 0000-0002-6222-630X, SPIN-код: 9930-5620, AuthorID: 587416, e-mail: iromanovarzn@bk.ru

THEORETICAL AND LEGAL ASPECTS OF THE PUNISHMENT SYSTEM OF ARTIFICIAL INTELLIGENCE

I.N. Romanova

Moscow University named after S. Witte, Ryazan

The purpose of the article is to study the issues of the need for the theoretical and legal development of a system of punishments of artificial intelligence in view of the fact that uncontrolled cybernetic creations carry potential and quite probable risks of the occurrence of undesirable consequences. Within the framework of the paper, the tasks of analyzing the social practice of using artificial intelligence technologies are solved; assessing risks and possible threats to the use of intelligent information systems; making proposals to prevent possible negative consequences of the practical implementation of intellectual digital resources, as well as leveling the undesirable consequences that have occurred; formulating and justifying the concept of punishments for the actions of artificial intelligence are solved. The methodological basis of the study was made up of general scientific methods of cognition, as well as the legal and technical private scientific method. The result of the conducted research is the formulation and justification of the position according to which the use of the system of punishments of artificial intelligence in any of its material manifestations, capable of acting as a coercive measure, seems appropriate. The author's conclusion based on the results of the consideration of the selected issues is the statement that overcoming possible obstacles in the mechanism of practical implementation of the system of punishments of artificial intelligence, which are not fundamentally insurmountable, directly depends on the efforts made.

Keywords: *probable criminological risks, artificial intelligence program, digital economy development, criminal liability, threat to public safety.*

About author:

ROMANOVA Irina – PhD in Law, Associate Professor, Head of the Department of Civil Law Disciplines, PEI HE «Moscow University named after S. Witte», Ryazan Branch, (390013, Ryazan, Pervomaisky av., 62), ORCID: 0000-0002-6222-630X, SPIN-code: 9930-5620, AuthorID: 587416, e-mail: iromanovarzn@bk.ru

Романова И.Н. Теоретико-правовые аспекты системы наказаний искусственного интеллекта // Вестник ТвГУ. Серия: Право. 2023. № 2 (74). С. 106–113.

Статья поступила в редакцию 06.12.2022 г.

Подписана в печать 27.05.2023 г.