

ПРОБЛЕМА «РАЗРЫВА ОТВЕТСТВЕННОСТИ» ЗА ВРЕД ПРИЧИНЕННЫЙ ИСПОЛЬЗОВАНИЕМ СИСТЕМ И ПРИЛОЖЕНИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Проблема ответственности за вред (ущерб), причиненный использованием систем и приложений искусственного интеллекта (далее – ИИ), активно обсуждается. В англоязычной юридической литературе для описания ситуации, когда затруднительно установить причинителя вреда и привлечь конкретных лиц к ответственности, используется термин *responsibility gap*, что в переводе обозначает «разрыв (пробел) ответственности». В российской правовой системе целесообразно применять аналогичный термин. В статье рассматривается специфика элементов юридического состава деликта, совершенного с использованием ИИ, определяется понятие «разрыв ответственности». Автор предлагает классифицировать виды «разрыва ответственности» в зависимости от технологических причин ошибок. Такая классификация необходима для последующей разработки механизма устранения «разрыва ответственности» юридическими средствами.

Ключевые слова: вред, ущерб, искусственный интеллект, разрыв ответственности, юридический состав, деликт

Для цитирования: Мельникова Е. Н. Проблема «разрыва ответственности» за вред причиненный использованием систем и приложений

¹ Мельникова Елена Николаевна – соискатель ученой степени кандидата юридических наук Санкт-Петербургского государственного университета, юрисконсульт отдела юридического сопровождения деятельности в сфере экономики и финансов Юридического управления Национального исследовательского университета ИТМО (Санкт-Петербург), ORCID: 0000-0001-7263-5281, e-mail: melnikova_elen5@mail.ru.

искусственного интеллекта // Российский юридический журнал. 2024. № 4. С. . DOI: 10.34076/20713797_2024_4_.

Введение

Проблема ответственности за вред (ущерб), причиненный использованием систем и приложений искусственного интеллекта (далее – ИИ), активно обсуждается в философской² и юридической литературе³ и по мере развития технологий становится все актуальнее.

В последние 10–15 лет в англоязычной юридической литературе появилось понятие *responsibility gap*, которое может быть переведено на русский язык и как «разрыв в ответственности», и как «пробел в

² Алексеев А. Ю. Философия искусственного интеллекта: концептуальный статус комплексного теста Тьюринга: автореф. дис. ... д-ра филос. наук. М., 2016. С. 4–5; Вислова А. Д. Искусственный интеллект как междисциплинарный феномен // Современная наука: актуальные проблемы теории и практики. Сер.: Познание. 2021. № 4. С. 84. DOI: 10.37882/2500-3682.2021.04.08; Sparrow R. Killer robots // Journal of Applied Philosophy. 2007. Vol. 24. № 1. P. 62. DOI: 10.1111/j.1468-5930.2007.00346.x; Simpson T. W., Müller V. C. Just war and robots' killings // The Philosophical Quarterly. 2016. Vol. 66. Is. 263. P. 302–303. DOI: 10.1093/pq/pqv075 и др.

³ См., например: Никитенко С. В. Концепции деликтной ответственности за вред, причиненный системами искусственного интеллекта // Вестник экономического правосудия Российской Федерации. 2023. № 1. С. 156–174. Доступ из СПС «КонсультантПлюс»; Гайворонская Я. В., Гальчун Е. А. Вред, причиненный искусственным интеллектом: аспекты ответственности и правосубъектности // Advances in Law Studies. 2021. Т. 9. № 4. URL: <https://naukaru.ru/ru/nauka/article/46754/view> (дата обращения: 23.01.2024); Померанец А. Д. Ответственность за вред, причиненный искусственным интеллектом // Цивилистика: право и процесс. 2021. № 4. С. 182; Balkin J. M. The Path of Robotics Law // California Law Review Circuit. 2015. Vol. 72. № 6. P. 49; Calo R. Robotics and the lessons of cyberlaw // California Law Review. 2015. Vol. 103. № 3. P. 542. DOI: <https://doi.org/10.2139/ssrn.2402972>; Pagallo U. The laws of robots: Crimes, contracts, and torts. Dordrecht: Springer, 2013; Robbins S. The many meanings of meaningful human control // AI Ethics. 2023. 24 July. DOI: 10.1007/s43681-023-00320-6. URL: <https://link.springer.com/article/10.1007/s43681-023-00320-6> (дата обращения: 02.10.2023).

ответственности». В русскоязычной юридической литературе данное понятие не применяется и не исследуется в контексте гражданско-правовой ответственности, а следовательно, не употребляется и обозначающий его термин. Однако специалисты отмечают, что «проблема привлечения к ответственности за вред, причиненный ИИ, является одной из наиболее сложных и обсуждаемых в научной литературе»⁴.

Одним из последствий отсутствия общепринятого терминологического обозначения проблемы является и ее содержательная неясность. Ответа требуют следующие вопросы: 1) что именно причиняет вред: использование системы ИИ, технологии ИИ или приложения ИИ – и в чем разница между ними; 2) в чем состоит специфика технологий ИИ, затрудняющая применение деликтной ответственности, а также поиск причинителя вреда; 3) что такое «разрыв ответственности»; 4) как дифференцируются юридические последствия в зависимости от технологических причин ошибок, возникших в ходе создания и использования ИИ?

«Разрыв ответственности» ведет к безнаказанности причинителя вреда и создает правовую неопределенность. Очевидно, что в условиях всеобщей цифровизации проблема «разрыва ответственности» проявляется особенно остро.

Определение понятий системы ИИ, технологии ИИ или приложения ИИ

В период формирования нормативно-правового регулирования разработки и использования ИИ важно выработать общепринятую терминологию.

⁴ Никитенко С. В. Международно-правовое регулирование использования искусственного интеллекта в области медицины: дис. ... канд. юрид. наук. СПб., 2023. С. 111.

В международной практике правового регулирования⁵ термином «система ИИ» (*artificial intelligence system*) принято обозначать программу для ЭВМ, компонентом которой является технология ИИ. В российском законодательстве термин «система ИИ» не употребляется. Наиболее близким понятием к западному термину является понятие «искусственный интеллект», впервые легально закрепленное в абз. «а» ст. 5 Национальной стратегией развития искусственного интеллекта на период до 2030 года, утвержденной Указом Президента РФ от 10 октября 2019 г. № 490 (ред. от 15 февраля 2024 г.) «О развитии искусственного интеллекта в Российской Федерации» (далее – Национальная стратегия). Из содержания понятия «ИИ» следует, что это также многокомпонентная программа для ЭВМ, поэтому в настоящей работе мы будем использовать термины «ИИ» и «система ИИ» как синонимы.

В российском законодательстве различаются понятия модели ИИ и технологии ИИ.

Согласно абз. «р» Национальной стратегии *модель ИИ* – это программа для электронных вычислительных машин (ее составная часть), предназначенная для выполнения интеллектуальных задач на уровне, сопоставимом с результатами интеллектуального труда человека или превосходящем их, использующая алгоритмы и наборы данных для выведения закономерностей, принятия решений или прогнозирования результатов. «Модель ИИ» – это легальный термин, введенный в российское законодательство Национальной

⁵ Artificial Intelligence Act European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)) P9_TA(2024)0138. URL: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf (дата обращения 31.01.2024); Draft Framework Convention on Artificial Intelligence, Human Rights, Democracy and The Rule of Law. Committee on Artificial Intelligence (CAI). Strasbourg, 18 December 2023. URL: <https://rm.coe.int/cai-2023-28-draft-framework-convention/1680ade043> (дата обращения 31.01.2024); Проект цифрового кодекса Кыргызской Республики. URL: <https://code.digital.gov.kg/ru/bills/6/> (дата обращения: 02.04.2024).

стратегией и обозначающий техническое явление, известное также как «модель машинного обучения».

Согласно абз. «б» Национальной стратегии под *технологиями ИИ* понимается совокупность технологий, включающая в себя компьютерное зрение, обработку естественного языка, распознавание и синтез речи, интеллектуальную поддержку принятия решений и перспективные методы искусственного интеллекта, что с технической точки зрения означает совокупность моделей машинного обучения.

Конечные продукты, в которых использованы технологии ИИ, в законодательстве РФ именуется такими терминами, как «решения, созданные с применением технологий ИИ»⁶, «продукты, которые созданы с использованием технологий ИИ»⁷. Подобные обозначения являются громоздкими, не отвечают предъявляемым к юридическим терминам требованиям⁸, таким как краткость, системность, четкость, однозначность, отсутствие синонимов. Соответственно,

⁶ Перечень поручений по итогам конференции «Путешествие в мир искусственного интеллекта» (утв. Президентом РФ 16 декабря 2021 г. № Пр-2371); постановление Правительства РФ от 9 февраля 2022 г. № 140 (ред. от 30 ноября 2022 г.) «О единой государственной информационной системе в сфере здравоохранения»; паспорт федерального проекта «Искусственный интеллект» национальной программы «Цифровая экономика Российской Федерации»; абз. «в» п. 19 Указа Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации» (вместе с Национальной стратегией развития искусственного интеллекта на период до 2030 года); приказ Минэкономразвития России от 29 июня 2021 г. № 391 «Об утверждении Порядка определения федеральным государственным бюджетным учреждением „Фонд содействия развитию малых форм предприятий в научно-технической сфере“ принадлежности проектов к проектам в сфере искусственного интеллекта».

⁷ Указ Президента РФ от 10 октября 2019 г. № 490 (ред. от 15 февраля 2024 г.) «О развитии искусственного интеллекта в Российской Федерации» (вместе с Национальной стратегией развития искусственного интеллекта на период до 2030 года).

⁸ *Толстик В. А.* Требования, предъявляемые к юридической терминологии: формально-логическое и социокультурное обоснование // *Юридическая техника.* 2016. № 10. С. 302–307.

любой продукт – устройство со встроенными технологиями ИИ целесообразно обозначать лаконичным термином «приложение ИИ», который мы будем использовать в данной работе.

Особенности технологий ИИ, затрудняющие поиск причинителя вреда

Для технологий ИИ не характерно принятие самостоятельных решений⁹, поэтому можно говорить о причинении вреда человеком в результате использования системы или приложения ИИ, но не о причинении вреда самой технологией или системой ИИ.

Технологии ИИ применяются в составе систем ИИ, что определяет главное отличие систем ИИ от других программ для ЭВМ. Технологии ИИ имеют существенную специфику.

Во-первых, работе всех без исключения технологий ИИ присуще получение вероятностного результата: модели машинного обучения производят вычисления, результатом которых является прогнозное число заданной точности; на его основе люди принимают решения, а системы ИИ совершают действия. Вероятностный характер этого числа – имманентное свойство технологий ИИ; назовем данное техническое явление «имманентной ошибкой модели ИИ». С точки зрения классических оснований юридической ответственности *вероятностный результат работы модели означает отсутствие причинно-следственной связи* между вредом и действиями лица, обязанного осуществлять эффективный контроль за использованием систем и приложений ИИ, тогда, когда происходит выдача ошибочного результата в пределах заданной погрешности модели.

Во-вторых, архитектура модели на стадии рабочего применения имеет динамический характер и может изменяться в ходе обучения¹⁰, которое

⁹ Melnikova E., Surov I. Legal. Status of Artificial Intelligence from Quantum-Theoretic Perspective // BRICS Law Journal. 2023. Vol. X. Is. 4. P. 22.

¹⁰ Мельникова Е. Н. Правовая квалификация модели машинного обучения на основе технического анализа этого процесса // Закон. 2023. № 6. С. 192.

происходит как с участием (контролируемое обучение), так и без участия человека (неконтролируемое обучение). При контролируемом обучении эффективный контроль за работой модели осуществляет лицо, которое ее фактически использует, производя либо адаптацию модели ИИ к новым данным, либо ее модификацию¹¹, что должно учитываться при распределении ответственности за неблагоприятные последствия использования «несамообучающейся» модели. Эффективный человеческий контроль за использованием «самообучающихся» моделей в автономных устройствах затруднен, что вызывает сложности с поиском причинителя вреда, ставшего результатом использования автономного устройства.

В-третьих, поскольку в ответ на получение прогнозного числа должно быть выполнено какое-либо действие, рабочее использование модели машинного обучения обычно происходит не изолированно (если только единственной целью модели не является получение прогнозного числа), а в составе программного комплекса (в том значении, которое ему придается в ГОСТ 19.101-77 «Виды программ и программных документов»), т. е. в составе системы ИИ, в которую в качестве компонентов может быть включено множество моделей (что в случае причинения вреда в результате сбоя затрудняет атрибутирование ошибки конкретному лицу).

Итак, технологии ИИ имеют ряд специфических черт, которые отличают их от обычных программ для ЭВМ и затрудняют поиск причинителя вреда. Среди них:

вероятностный характер результата работы с имманентно присущими ИИ ошибками, в которых никто не виноват;

динамические свойства программной архитектуры, что обуславливает возможность изменения свойств модели в процессе ее жизненного цикла;

рабочее использование модели машинного обучения в составе программного или аппаратно-программного комплекса, что осложняет поиск

¹¹ Там же. С. 193.

компонента, в котором произошел сбой, повлекший неблагоприятные последствия.

Перечисленные свойства технологий ИИ являются причиной «разрыва подотчетности».

Проблема технологически размытой подотчетности возникла еще до появления ИИ: «разрыв подотчетности» обусловлен сложностью тех или иных программ или устройств, поэтому пробелы в ответственности могут возникать и в необучаемых системах, не имеющих отношения к ИИ¹². Однако с появлением ИИ проблема «разрыва подотчетности» и «разрыва ответственности» встала особенно остро.

Пробел подотчетности (*accountability gap*) является следствием «размывания» сфер технологического контроля, обусловленного технологическими особенностями программы или устройства. *Подотчетность – это прежде всего технологический феномен.* Категория вины должна быть применена к надлежащему субъекту, поиск которого лежит в области технической экспертизы. Когда техническая экспертиза не может установить причинителя вреда, возникает *правовое явление*, которое получило название «разрыв ответственности» (*responsibility gap*). Этот термин в последнее время устойчиво ассоциируется с проблемой возложения ответственности за вред, причиненный ИИ.

Понятие «разрыва ответственности»

Классическое и самое общее правило возложения любой, включая гражданско-правовую, ответственности гласит: «Агент может считаться ответственным только в том случае, если он знает конкретные факты, связанные с его действиями, и если он способен свободно принять решение действовать и выбрать одно из доступных альтернативных действий,

¹² *Santoni de Sio F., Mecacci G.* Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them // *Philosophy & Technology*. 2021. Vol. 34. P. 1057–1084. DOI: 10.1007/s13347-021-00450-x. URL: <https://link.springer.com/article/10.1007/s13347-021-00450-x> (дата обращения: 02.10.2023).

основанных на этих фактах»¹³. Однако системы и приложения ИИ создают новую ситуацию, когда производитель или оператор машины оказывается не способен предсказывать ее будущее поведение, поэтому возникает правовая проблема возложения юридической ответственности за вред, причиненный использованием систем и приложений ИИ, на конкретное лицо или лиц тогда, когда противоправность действий отсутствует. В результате общество «сталкивается с разрывом в ответственности, который не может быть преодолен с помощью традиционных концепций возложения ответственности»¹⁴.

В российской юридической доктрине понятия «разрыв ответственности», «подотчетность» не имеют широкого применения. В европейской юридической литературе термин *responsibility gap* наиболее часто употребляется при анализе юридической ответственности за неблагоприятные последствия использования автономных летальных систем вооружения, хотя встречаются работы, где данные понятия используются в гражданско-правовом смысле¹⁵.

Проблема «разрыва ответственности» относительно давно обсуждается в рамках концепций ответственности в философии (теории философской, юридической и социологической ответственности)¹⁶ и в рамках дискуссии об

¹³ *Matthias A.* The responsibility gap: Ascribing responsibility for the actions of learning automata // *Ethics and Information Technology*. 2004. Vol. 6. № 3. P. 175. DOI: 10.1007/s10676-004-3422-1.

¹⁴ *Ibid.* P. 175.

¹⁵ *Santoni de Sio F., Mecacci G.* *Op. cit.*; *Matthias A.* *Op. cit.* P. 175.

¹⁶ *Bovens M.* The quest for responsibility: Accountability and citizenship in complex organisations. Cambridge; N. Y.: Cambridge University Press, 1998. 254 p. DOI: 10.2307/2667065. URL: https://www.researchgate.net/publication/275840305_The_Quest_for_Responsibility_Accountability_and_Citizenship_in_Complex_Organisations (дата обращения: 02.12.2023); *Collingridge D.* The Social Control of Technology. L.: Frances Pinter, 1980. 200 p. DOI: 10.2307/2634327; *Danaher J.* Robots, law and the retribution gap // *Ethics and Information Technology*. 2016. Vol. 18. P. 299–309. DOI: 10.1007/s10676-016-9403-3; *Pesch U.* Engineers and Active Responsibility //

объяснимости алгоритмов и искусственного интеллекта¹⁷. Однако наиболее четко «разрыв ответственности» как юридическую проблему описал А. Маттиас в 2004 г.¹⁸ В качестве основной причины возникновения «разрывов ответственности» Маттиас выделил способность автономных систем ИИ к обучению.

Существование проблемы «разрыва ответственности» применительно к неблагоприятным последствиям использования ИИ признается не всеми специалистами. В англоязычной юридической литературе можно выделить три основных подхода к этому вопросу. Сторонники подхода, названного Ф. Антонио де Сियो и Дж. Мекаччи «фатализмом», представляют проблему «разрыва ответственности» как новую и неразрешимую¹⁹. Приверженцы подхода, получившего название «дефационизм», отвергают проблему «разрыва в ответственности» как ложную²⁰. К представителям «дефационизма» можно отнести тех специалистов, кто придерживается мнения о достаточности имеющегося правового регулирования ИИ²¹. Подход, согласно которому «разрыв ответственности» признается и предлагается устранить внедрением

Science and Engineering Ethics. 2015. Vol. 21. № 4. P. 925–939. DOI: 10.1007/s11948-014-9571-7; Van de Poel I., Sand M. Varieties of responsibility: two problems of responsible innovation // Synthese. 2018. Vol. 198. P. 4769–4787. DOI: 10.1007/s11229-018-01951-7.

¹⁷ Mittelstadt B. D., Allo P., Taddeo M., Wachter S., Floridi L. The ethics of algorithms: Mapping the debate // Big Data & Society. 2016. Vol. 3. № 2. P. 1–21. DOI: 10.1177/2053951716679679; Doran D., Schulz S., Besold T. R. What does explainable ai really mean? A new conceptualization of perspectives // ArXiv. 2017. DOI: 10.48550/arXiv.1710.00794.

¹⁸ Matthias A. Op. cit. P. 175.

¹⁹ Sparrow R. Op. cit.; Matthias A. Op. cit. P. 175.

²⁰ Simpson T. W., Müller V. C. Op. cit. P. 305.

²¹ Burri T. International Law and Artificial Intelligence // German Yearbook of International Law 2017. Berlin: Duncker & Humblot, 2019. Vol. 60. P. 101; Hill S., Marsan N. Artificial Intelligence and Accountability: A Multinational Legal Perspective // NATO Office of Legal Affairs Report STO-MP-IST-160. 2018. 25 May. P. 8. URL: <https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-IST-160/MP-IST-160-PP-4P.pdf> (дата обращения: 24.07.2023).

новых технических и (или) юридических инструментов, Ф. Антонио де Сियो и Дж. Мекаччи отнесли к «решенизму», обозначив два варианта устранения «разрыва ответственности» - технический и юридический²².

Общепринятого определения «разрыва ответственности» в контексте использования ИИ не существует. Однако ввиду назревающей необходимости формирования нормативно-правовой базы распределения ответственности за вред, причиненный ИИ, думается, термин «разрыв ответственности» может использоваться для обозначения правового явления, характеризующегося затруднениями в поиске причинителя вреда и привлечении к ответственности конкретных лиц. Понятие, обозначаемое этим термином, должно получить четкое доктринальное определение, а его причины должны быть классифицированы для облегчения поиска способов устранения «разрыва ответственности» юридическими средствами.

Проблема формирования юридического состава деликта, совершенного с использованием ИИ

Юридический состав *генерального деликта* включает в себя следующие элементы: 1) наличие вреда; 2) противоправный характер поведения лица; 3) причинно-следственную связь между действиями причинителя вреда и вредом; 4) вину причинителя вреда.

Когда использование технологий ИИ приводит к тому, что ошибку нельзя атрибутировать конкретному лицу, оказывается невозможным установить следующие элементы юридического состава деликта: противоправный характер поведения лица (лиц) тогда, когда, ни один из участников жизненного цикла системы или приложения ИИ не нарушил закон и инструкции; причинно-следственную связь между действиями причинителя вреда и вредом; вину. В случае причинения вреда системой или приложением ИИ однозначно можно констатировать только факт причинения вреда. Основными проблемами привлечения к ответственности за вред, причиненный ИИ, является поиск причинителя вреда и отсутствие его вины тогда, когда он не мог повлиять на

²² *Santoni de Sio F., Mecacci G. Op. cit. P. 1072*

неблагоприятные последствия использования ИИ иначе, как отказавшись от такого использования.

Охранительные обязательства, возникающие в случае причинения вреда использованием системы или приложения ИИ, относятся к группе деликтных обязательств, характеризующихся особенностями средств причинения вреда. Сегодня известны следующие виды *специальных деликтов*: а) причинение вреда деятельностью, создающей повышенную опасность для окружающих; б) причинение вреда вследствие недостатков товаров, работ, услуг²³. Попробуем «примерить» конструкции специальных деликтов к охранительным обязательствам из причинения вреда ИИ.

1. Ответственность владельца источника повышенной опасности (ИПО). Одним из наиболее известных приложений ИИ, обладающих признаками ИПО, сегодня является высокоавтономное транспортное средство (ВАТС). Однако очевидно, что скоро появятся и другие приложения ИИ, в том числе полностью автономные, например роботы-помощники с широким функционалом. Основным аргументом противников отнесения ВАТС к ИПО является то, что владелец изначально не может непосредственно осуществлять контроль за транспортным средством, однако «владение источником повышенной опасности, в силу ст. 1079 ГК РФ, предполагает не только физическое обладание им, но и его эксплуатацию, и управление им, особенно если этот источник является транспортным средством»²⁴. В некоторых приложениях ИИ, например в беспилотном транспортном средстве, у водителя нет руля и педалей²⁵, что априори исключает возможность человеческого контроля.

²³ Шевченко А. С., Шевченко Г. Н. Деликтные обязательства в российском гражданском праве: учеб. пособие. М.: Статут. 2013. Доступ из СПС «КонсультантПлюс».

²⁴ Ли Ч., Менг С. Внедоговорная ответственность за вред, причиненный беспилотными транспортными средствами // Закон. 2020. № 3. С. 49–56. Доступ из СПС «КонсультантПлюс».

²⁵ Там же.

Конструкция ИПО предполагает ответственность владельца без вины в случае потери контроля за вещью, но не ответственность за чужие непредотвратимые ошибки, например за ошибки разработчика или имманентную ошибку модели. Поэтому презумпция вины оказывается теоретически не обоснованной. Кроме того, состав данного специального деликта не исключает обязательность установления причинно-следственной связи и противоправности действий владельца (если владелец ВАТС соблюдал ПДД и инструкцию производителя ВАТС, то противоправность его действий отсутствует). Соответственно, возложение ответственности за чужие действия (бездействие) на владельца ИПО в рамках данной конструкции само по себе невозможно: не хватает юридических инструментов для восполнения как минимум двух недостающих элементов состава.

2. Ответственность за вред, причиненный вследствие недостатков товаров. В соответствии со ст. 1095 ГК РФ вред, причиненный вследствие конструктивных, рецептурных или иных недостатков товара, а также вследствие недостаточной или недостоверной информации о товаре, подлежит возмещению продавцом или изготовителем товара независимо от их вины и от того, состоял потерпевший с ними в договорных отношениях или нет.

В этом случае проблемным элементом состава является наличие причинно-следственной связи между дефектом (ошибкой) системы или приложения ИИ и вредом, поскольку: а) понятие дефекта (ошибки) легально не закреплено в законодательстве РФ; б) свойства системы или приложения ИИ могут изменяться под воздействием пользователя. Соответственно, для данной конструкции проблематично установление двух элементов состава деликта.

«Разрыв ответственности» определяется отсутствием возможности установления необходимых элементов юридического состава деликта, совершенного с использованием ИИ. «Разрыв ответственности» крайне нежелателен, поэтому следует найти возможности для его устранения. Для возложения ответственности на конкретное лицо (лиц) необходима разработка

юридических инструментов для восполнения недостающих элементов состава деликта, совершенного с использованием ИИ.

Виды «разрыва ответственности»

Как уже было сказано, «разрыв ответственности» является одним из правовых последствий технологически размытой подотчетности. Однако «разрывы подотчетности» могут порождаться разными технологическими причинами.

Технологический «разрыв подотчетности» возникает в случае причинения вреда использованием ИИ как минимум в двух ситуациях:

1) имманентная ошибка модели ИИ объективно непредотвратима (например, ошибка распознавания предмета при достаточном обучении беспилотника);

2) имманентная ошибка модели ИИ исключена экспертизой, но поиск «автора» ошибки существенно затруднен в связи со сложностью системы или приложения ИИ, в которых произошел сбой (проблема «черного ящика»)²⁶.

С юридической точки зрения имманентная ошибка модели ИИ означает отсутствие возможности установления трех элементов юридического состава гражданско-правового нарушения: а) причинно-следственной связи между вредом и действиями лица, использующего систему или приложение ИИ; б) вины этого лица; в) противоправности. Это явление, как представляется, наиболее близко к тому, что в литературе принято называть «разрывом ответственности»²⁷. Думается, только в этом случае можно говорить об истинном «разрыве подотчетности», результатом которого выступает *истинный «разрыв ответственности»* – юридическое явление, при котором

²⁶ Наумов В. Б. Отказ от цифровых технологий: абсурд или новое право человека и гражданина // Четвертые Бачиловские чтения: материалы Междунар. науч.-практ. конф. М.: Ин-т государства и права РАН, 2022. С. 78.

²⁷ Königs P. Artificial intelligence and responsibility gaps: what is the problem? // Ethics and Information Technology. 2022. Vol. 24. № 3. DOI: 10.1007/s10676-022-09643-0; Santoni de Sio F., Mecacci G. Op. cit.

объективно невозможно определить причинителя вреда ввиду имманентной ошибки модели ИИ. Такой вид «разрыва ответственности» может быть преодолен только юридическими средствами, направленными на восполнение трех недостающих элементов состава.

Когда техническая экспертиза установила отсутствие имманентной ошибки модели ИИ, но все еще не в состоянии ответить, чьи действия привели к ошибке (ситуация «черного ящика»), как представляется, имеет место мнимый «разрыв подотчетности», из которого вытекает *мнимый «разрыв ответственности»*. Мы назвали «разрыв ответственности» мнимым, поскольку хотя при текущем уровне техники имеется субъективная невозможность выявления «автора» ошибки, но в будущем такая возможность может появиться и тогда причинно-следственная связь может быть установлена (а вместе с ней и другие элементы состава). В такой ситуации мнимый «разрыв ответственности» будет устранен, по сути, техническими средствами. Однако с учетом сегодняшнего уровня техники мнимый «разрыв ответственности» может быть преодолен только юридическими средствами, как и в первой ситуации. Думается, что юридические средства устранения истинного и мнимого «разрывов ответственности» должны быть концептуально схожими.

«Разрыв ответственности» (как истинный, так и мнимый) необходимо отличать от «размывания ответственности». «Размывание ответственности» – ситуация, когда поиск причинителя вреда затруднен в связи с так называемой проблемой «многих рук», при которой работа системы ИИ, управляющей приложением ИИ, опосредуется действиями различных лиц: производителя, владельца, диспетчера, оператора, лица, отвечающего за техническое обслуживание, владельца облачной инфраструктуры и иных участников жизненного цикла приложения ИИ. В такой ситуации поиск причинителя вреда определяется качеством и глубиной технологической экспертизы, которая должна установить лиц, чьи действия привели к ошибке, повлекшей неблагоприятные последствия использования приложения ИИ. В этом случае состав деликта установить *возможно*. Но если глубины экспертизы не хватает,

то такие лица объективно не могут быть установлены. И тогда, как представляется, имеет место не «размывание», а «разрыв ответственности» с правовыми последствиями, характерными для двух ситуаций, рассмотренных выше.

Предложенная классификация необходима для последующей разработки механизма устранения «разрыва ответственности» юридическими средствами, которые должны учитывать объективную и субъективную невозможность атрибутирования ошибки.

При выборе юридических средств, направленных на устранение «разрыва ответственности», «следует с осторожностью относиться к конструкциям усеченных составов деликтов, которые исключают вину как условие привлечения к ответственности и возлагают ответственность на одну группу лиц»²⁸ (к примеру, пользователей²⁹). Для устранения «разрыва ответственности» необходимо такое восполнение недостающих элементов состава гражданско-правового деликта юридическими средствами, которое позволит справедливо распределить ответственность за вред, причиненный ИИ, между участниками жизненного цикла систем и приложений ИИ. Выбор этих юридических средств зависит от вида «разрыва ответственности» и требует разработки нового юридического подхода, позволяющего преодолеть недостатки конструкций усеченных составов деликтов. Думается, что такими юридическими средствами могут быть презумпции и фикции (которые сами по себе не новы), выработанные в соответствии с концепцией значимого человеческого контроля за использованием ИИ и представляющие собой механизм поиска лица, ответственного за вред, причиненный ИИ.

Заключение

²⁸ *Никитенко С. В.* Международно-правовое регулирование использования искусственного интеллекта в области медицины. С. 115–116.

²⁹ *Буйлов М.* ИИ: безответственный, но доходный // *Kommersant.ru*. 2023. 20 июля. URL: <https://www.kommersant.ru/doc/6111440> (дата обращения: 20.07.2023).

Предлагаем определить «разрыв ответственности» за вред, причиненный использованием ИИ, как правовое явление, при котором возложение ответственности за причиненный вред на конкретное лицо (лиц) согласно существующим основаниям деликтной ответственности невозможно ввиду невозможности установления таких элементов состава гражданско-правового нарушения, как причинно-следственная связь между вредом и действиями конкретных лиц, противоправность и виновность этих действий.

«Разрывы ответственности» за вред (ущерб), причиненный использованием ИИ, можно разделить на три вида:

истинный пробел ответственности – объективная невозможность определить причинителя вреда ввиду возникновения неблагоприятных последствий имманентной ошибки модели ИИ, действующей в составе системы ИИ (такая ошибка обусловлена вероятностной природой выдачи результата и поэтому не зависит от контролирующего субъекта);

мнимый пробел ответственности – субъективная невозможность определить причинителя вреда ввиду возникновения неблагоприятных последствий ошибки системы ИИ, когда выводы технической экспертизы не позволяют вменить вину конкретному лицу (лицам) ввиду неясной этиологии ошибки (проблема «черного ящика»), которая не является имманентной ошибкой конкретной модели ИИ;

«размывание ответственности» – субъективная невозможность определить причинителя вреда ввиду возникновения неблагоприятных последствий использования приложения ИИ, когда работа управляющей им системы ИИ опосредуется действиями многих участников жизненного цикла приложения ИИ.

Для устранения «разрыва ответственности» за вред, причиненный ИИ, необходима разработка нового юридического подхода к поиску причинителя вреда в случае невозможности атрибутирования ошибки посредством экспертизы.

Список литературы

Алексеев А. Ю. Философия искусственного интеллекта: концептуальный статус комплексного теста Тьюринга: автореф. дис. ... д-ра филос. наук. М., 2016. 44 с.

Буйлов М. ИИ: безответственный, но доходный // *Kommersant.ru*. 2023. 20 июля. URL: <https://www.kommersant.ru/doc/6111440> (дата обращения: 20.07.2023).

Вислова А. Д. Искусственный интеллект как междисциплинарный феномен // *Современная наука: актуальные проблемы теории и практики*. Серия: Познание. 2021. № 4. С. 84–88. DOI: 10.37882/2500-3682.2021.04.08.

Гайворонская Я. В., Гальчун Е. А. Вред, причиненный искусственным интеллектом: аспекты ответственности и правосубъектности // *Advances in Law Studies*. 2021. Т. 9. № 4. URL: [https://https://naukaru.ru/ru/nauka/article/46754/view](https://naukaru.ru/ru/nauka/article/46754/view) (дата обращения: 23.01.2024).

Ли Ч., Менг С. Внедоговорная ответственность за вред, причиненный беспилотными транспортными средствами // *Закон*. 2020. № 3. С. 49–56.

Мельникова Е. Н. Правовая квалификация модели машинного обучения на основе технического анализа этого процесса // *Закон*. 2023. № 6. С. 187–196.

Наумов В. Б. Отказ от цифровых технологий: абсурд или новое право человека и гражданина // *Четвертые Бачиловские чтения: материалы Междунар. науч.-практ. конф.* М.: Ин-т государства и права РАН, 2022. С. 78–84.

Никитенко С. В. Концепции деликтной ответственности за вред, причиненный системами искусственного интеллекта // *Вестник экономического правосудия Российской Федерации*. 2023. № 1. С. 156–174.

Никитенко С. В. Международно-правовое регулирование использования искусственного интеллекта в области медицины: дис. ... канд. юрид. наук. СПб., 2023. 223 с.

Померанец А. Д. Ответственность за вред, причиненный искусственным интеллектом // *Цивилистика: право и процесс*. 2021. № 4. С. 181–188.

Толстик В. А. Требования, предъявляемые к юридической терминологии: формально-логическое и социокультурное обоснование // Юридическая техника. 2016. № 10. С. 302–307.

Шевченко А. С., Шевченко Г. Н. Деликтные обязательства в российском гражданском праве: учеб. пособие. М.: Статут. 2013. 133 с.

Balkin J. M. The Path of Robotics Law // California Law Review Circuit. 2015. Vol. 72. № 6. P. 45–60.

Bovens M. The quest for responsibility: Accountability and citizenship in complex organisations. Cambridge; N. Y.: Cambridge University Press, 1998. 254 p. DOI: 10.2307/2667065. URL: https://www.researchgate.net/publication/275840305_The_Quest_for_Responsibility_Accountability_and_Citizenship_in_Complex_Organisations (дата обращения: 02.12.2023).

Burri T. International Law and Artificial Intelligence // German Yearbook of International Law 2017. Berlin: Duncker & Humblot, 2019. Vol. 60. P. 91–108.

Calo R. Robotics and the lessons of cyberlaw // California Law Review. 2015. Vol. 103. № 3. P. 513–563. DOI: 10.2139/ssrn.2402972.

Collingridge D. The Social Control of Technology. L.: Frances Pinter, 1980. 200 p. DOI: 10.2307/2634327.

Danaher J. Robots, law and the retribution gap // Ethics and Information Technology. 2016. Vol. 18. P. 299–309. DOI: 10.1007/s10676-016-9403-3.

Doran D., Schulz S., Besold T. R. What does explainable ai really mean? A new conceptualization of perspectives // Harvard University Press. 2017. DOI: 10.48550/arXiv.1710.00794.

Hill S., Marsan N. Artificial Intelligence and Accountability: A Multinational Legal Perspective // NATO Office of Legal Affairs Report STO-MP-IST-160. 2018. 25 May. URL: <https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-IST-160/MP-IST-160-PP-4P.pdf> (дата обращения: 24.07.2023).

Königs P. Artificial intelligence and responsibility gaps: what is the problem? // *Ethics and Information Technology*. 2022. Vol. 24. № 3. DOI: 10.1007/s10676-022-09643-0.

Matthias A. The responsibility gap: Ascribing responsibility for the actions of learning automata // *Ethics and Information Technology*. 2004. Vol. 6. № 3. P. 175–183. DOI: 10.1007/s10676-004-3422-1.

Melnikova E., Surov I. Legal. Status of Artificial Intelligence from Quantum-Theoretic Perspective // *BRICS Law Journal*. 2023. Vol. X. Is. 4. P. 5–34.

Mittelstadt B. D., Allo P., Taddeo M., Wachter S., Floridi L. The ethics of algorithms: Mapping the debate // *Big Data & Society*. 2016. Vol. 3. № 2. P. 1–21. DOI: 10.1177/2053951716679679.

Pagallo U. The laws of robots: Crimes, contracts, and torts. Dordrecht: Springer, 2013. 200 p.

Pesch U. Engineers and Active Responsibility // *Science and Engineering Ethics*. 2015. Vol. 21. № 4. P. 925–939. DOI: 10.1007/s11948-014-9571-7.

Robbins S. The many meanings of meaningful human control // *AI Ethics*. 2023. 24 July. DOI: 10.1007/s43681-023-00320-6. URL: <https://link.springer.com/article/10.1007/s43681-023-00320-6> (дата обращения: 02.10.2023).

Santoni de Sio F., Mecacci G. Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them // *Philosophy & Technology*. 2021. Vol. 34. P. 1057–1084 DOI: 10.1007/s13347-021-00450-x. URL: <https://link.springer.com/article/10.1007/s13347-021-00450-x> (дата обращения: 02.10.2023).

Simpson T. W., Müller V. C. Just war and robots' killings // *The Philosophical Quarterly*. 2016. Vol. 66. Is. 263. P. 302–322. DOI: 10.1093/pq/pqv075.

Sparrow R. Killer robots // *Journal of Applied Philosophy* 2007. Vol. 24. № 1. P. 62–77. DOI: 10.1111/j.1468-5930.2007.00346.x.

Van de Poel I., Sand M. Varieties of responsibility: two problems of responsible innovation // Synthese. 2018. Vol. 198. P. 4769–4787. DOI: 10.1007/s11229-018-01951-7.

UDC 347

DOI: 10.34076/20713797_2024_4_

THE PROBLEM OF THE «RESPONSIBILITY GAP» FOR HARM CAUSED BY ARTIFICIAL INTELLIGENCE SYSTEMS AND APPLICATIONS

Melnikova Elena – master of science, legal counsel, National Research University ITMO (St. Petersburg), ORCID: 0000-0001-7263-5281, e-mail: melnikova_elen5@mail.ru.

The issue of responsibility for harm (damage) caused by the use of artificial intelligence (AI) systems and applications is a topic of active discussion. In English-language legal literature, the concept of «responsibility gap» is used to describe situations where it is challenging to identify the cause of harm and hold specific individuals accountable. It would be beneficial to introduce a similar term into the Russian legal framework. The article explores the characteristics of the elements constituting a tort committed through the use of AI, and defines the concept of responsibility gap. The author proposes classifying responsibility gaps based on the technological factors that lead to errors. This classification is essential for developing a mechanism to address responsibility gaps through legal means in the future.

Key words: harm, damage, artificial intelligence, responsibility gap, legal composition, tort

For citation: Melnikova E. (2024) The problem of the «responsibility gap» for harm caused by artificial intelligence systems and applications. In *Rossiiskii yuridicheskii zhurnal*, no. 4, pp. , DOI: 10.34076/20713797_2024_4_.

References

Alekseev A. Yu. (2016) *Filosofiya iskusstvennogo intellekta: kontseptual'nyi status kompleksnogo testa T'yuringa: avtoref. dis. ... d-ra filos. nauk* [Philosophy of

artificial intelligence: conceptual status of the complex Turing test: a candidate of philosophical sciences thesis]. Moscow, 44 p.

Balkin J. M. (2015) The Path of Robotics Law. In *California Law Review Circuit*. Vol. 72. no. 6. pp. 45–60.

Bovens M. (1998) *The quest for responsibility: Accountability and citizenship in complex organisations*. Cambridge, New York, Cambridge University Press, 254 p., DOI: 10.2307/2667065, available at: https://www.researchgate.net/publication/275840305_The_Quest_for_Responsibility_Accountability_and_Citizenship_in_Complex_Organisations (accessed: 02.12.2023).

Builov M. (2023) II: bezotvetstvennyi, no dokhodnyi [AI: Irresponsible, but Profitable]. In *Kommersant.ru*, 20 July, available at: <https://www.kommersant.ru/doc/6111440> (accessed: 20.07.2023).

Burri T. (2019) International Law and Artificial Intelligence. In *German Yearbook of International Law 2017*. Berlin, Duncker & Humblot, vol. 60, pp. 91–108.

Calo R. (2015) Robotics and the lessons of cyberlaw. In *California Law Review*, vol. 103, no. 3, pp. 513–563, DOI: 10.2139/ssrn.2402972.

Collingridge D. (1980) *The Social Control of Technology*. London, Frances Pinter, 200 p., DOI: 10.2307/2634327.

Danaher J. (2016) Robots, law and the retribution gap. In *Ethics and Information Technology*, vol. 18, pp. 299–309, DOI: 10.1007/s10676-016-9403-3.

Doran D., Schulz S., Besold T. R. (2017) What does explainable AI really mean? A new conceptualization of perspectives. In *ArXiv*, DOI: 10.48550/arXiv.1710.00794.

Gaivoronskaya Ya. V., Gal'chun E. A. (2021) Vred, prichinennyi iskusstvennym intellektom: aspekty otvetstvennosti i pravosub'ektnosti [Harm caused by artificial intelligence: aspects of liability and legal capacity]. In *Advances in Law Studies*, vol. 9, no. 4, available at: <https://naukaru.ru/ru/nauka/article/46754/view> (accessed: 23.01.2024).

Hill S., Marsan N. (2018) Artificial Intelligence and Accountability: A Multinational Legal Perspective. In *NATO Office of Legal Affairs Report STO-MP-IST-160*, 25 May, available at: <https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-IST-160/MP-IST-160-PP-4P.pdf> (accessed: 24.07.2023).

Königs P. (2022) Artificial intelligence and responsibility gaps: what is the problem? In *Ethics and Information Technology*, vol. 24, no. 3, DOI: 10.1007/s10676-022-09643-0.

Li Ch., Meng S. (2020) Vnedogovornaya otvetstvennost' za vred, prichinennyi bespilotnymi transportnymi sredstvami [Non-contractual liability for harm caused by unmanned vehicles]. In *Zakon*, no. 3, pp. 49–56.

Matthias A. (2004) The responsibility gap: Ascribing responsibility for the actions of learning automata. In *Ethics and Information Technology*, vol. 6, no. 3, pp. 175–183, DOI: 10.1007/s10676-004-3422-1.

Mel'nikova E. N. (2023) Pravovaya kvalifikatsiya modeli mashinnogo obucheniya na osnove tekhnicheskogo analiza etogo protsessa [Legal qualification of a machine learning model based on a technical analysis of this process]. In *Zakon*, no. 6, pp. 187–196.

Melnikova E., Surov I. (2023) Legal. Status of Artificial Intelligence from Quantum-Theoretic Perspective. In *BRICS Law Journal*, vol. X, is. 4, pp. 5–34.

Mittelstadt B. D., Allo P., Taddeo M., Wachter S., Floridi L. (2016) The ethics of algorithms: Mapping the debate. In *Big Data & Society*, vol. 3, no. 2, pp. 1–21. DOI: 10.1177/2053951716679679.

Naumov V. B. (2022) Otkaz ot tsifrovyykh tekhnologii: absurd ili novoe pravo cheloveka i grazhdanina [Refusal of digital technologies: absurdity or a new right of man and citizen]. In *Chetvertye Bachilovskie chteniya: conference papers*. Moscow, Institut gosudarstva i prava RAN, pp. 78–84.

Nikitenko S. V. (2023) Kontseptsii deliktnoi otvetstvennosti za vred, prichinennyi sistemami iskusstvennogo intellekta [Concepts of tort liability for harm

caused by artificial intelligence systems]. In *Vestnik ekonomicheskogo pravosudiviya Rossiiskoi Federatsii*, no. 1, pp. 156–174.

Nikitenko S. V. (2023) *Mezhdunarodno-pravovoe regulirovanie ispol'zovaniya iskusstvennogo intellekta v oblasti meditsiny: dis. ... kand. yurid. nauk* [International legal regulation of the use of artificial intelligence in the field of medicine: a candidate of legal sciences thesis]. Saint Petersburg, 223 p.

Pagallo U. (2013) *The laws of robots: Crimes, contracts, and torts*. Dordrecht, Springer, 200 p.

Pesch U. (2015) Engineers and Active Responsibility. In *Science and Engineering Ethics*, vol. 21, no. 4, pp. 925–939, DOI: 10.1007/s11948-014-9571-7.

Pomeranets A. D. (2021) Otvetstvennost' za vred, prichinennyi iskusstvennym intellektom [Liability for harm caused by artificial intelligence]. In *Tsivilistika: pravo i protsess*, no. 4, pp. 181–188.

Robbins S. (2023) The many meanings of meaningful human control. In *AI Ethics*, 24 July, DOI: 10.1007/s43681-023-00320-6, available at: <https://link.springer.com/article/10.1007/s43681-023-00320-6> (accessed: 02.10.2023).

Santoni de Sio F., Mecacci G. (2021) Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them. In *Philosophy & Technology*, vol. 34, pp. 1057–1084, DOI: 10.1007/s13347-021-00450-x, available at: <https://link.springer.com/article/10.1007/s13347-021-00450-x> (accessed: 02.10.2023).

Shevchenko A. S., Shevchenko G. N. (2013) *Deliktnye obyazatel'stva v rossiiskom grazhdanskom prave* [Tort obligations in Russian civil law]. Moscow, Statut, 133 p.

Simpson T. W., Müller V. C. (2016) Just war and robots' killings. In *The Philosophical Quarterly*, vol. 66, is. 263, pp. 302–322, DOI: 10.1093/pq/pqv075.

Sparrow R. (2007) Killer robots. In *Journal of Applied Philosophy*, vol. 24, no. 1, pp. 62–77, DOI: 10.1111/j.1468-5930.2007.00346.x.

Tolstik V. A. (2016) Trebovaniya, pred'yavlyaemye k yuridicheskoi terminologii: formal'no-logicheskoe i sotsiokul'turnoe obosnovanie [Requirements for legal terminology: formal-logical and socio-cultural justification]. In *Yuridicheskaya tekhnika*, no. 10, pp. 302–307.

Van de Poel I., Sand M. (2018) Varieties of responsibility: two problems of responsible innovation. In *Synthese*, vol. 198, pp. 4769–4787, DOI: 10.1007/s11229-018-01951-7.

Vislova A. D. (2021) Iskusstvennyi intellekt kak mezhdistsiplinarnyi fenomen [Artificial intelligence as an interdisciplinary phenomenon]. In *Sovremennaya nauka: aktual'nye problemy teorii i praktiki. Ser.: Poznanie*, no. 4, pp. 84–88, DOI: 10.37882/2500-3682.2021.04.08.