DOI: https://doi.org/10.24833/0869-0049-2024-4-132-145

Исследовательская статья
УДК: 341.018
Поступила в редакцию: 19.06.2024
Принята к публикации: 14.11.2024

Елена Николаевна МЕЛЬНИКОВА
Санкт-Петербургский государственный университет
Университетская наб., д. 7–9, Санкт-Петербург, 199034, Российская Федерация
melnikova_elena5@mail.ru
ORCID: 0000-0001-7263-5281

ПРИНЦИП ОТВЕТСТВЕННОСТИ КОНТРОЛИРУЮЩЕГО ЛИЦА КАК ОСНОВА ДЛЯ УСТРАНЕНИЯ «РАЗРЫВА ОТВЕТСТВЕННОСТИ» ЗА ВРЕД, ПРИЧИНЕННЫЙ ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ

ВВЕДЕНИЕ. *Использование технологий искусственного интеллекта (далее – ИИ) характеризуется опосредованием действий человека автономными процессами, что приводит в случае, когда техническая экспертиза не в состоянии выявить причинителя вреда, к «разрыву ответственности» – нежелательному правовому явлению, при котором возложение ответственности за вред, причиненный использованием ИИ, на конкретное лицо (лиц) по правилам деликтной ответственности невозможно.*

МАТЕРИАЛЫ И МЕТОДЫ. *При проведении исследования использовались общенаучные и специальные методы, в том числе исторический метод, методы формальной логики, анализа, синтеза, а также системные и сравнительно-правовые методы.*

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ. *Для устранения «разрыва ответственности» в статье предложен механизм, позволяющий восполнить недостающие элементы состава деликта, совершенного с использованием ИИ тогда, когда ошибку, которая привела к причинению вреда, de lege lata невозможно атрибутировать ни одному участнику жизненного цикла системы или приложения ИИ. Отправной точкой разработки данного механизма стала теория «руководящего контроля» за использованием ИИ. Юридическое осмысление философских оснований теории «руководящего контроля» позволяет обосновать общеправовой принцип распределения ответственности за вред, причиненный ИИ, согласно которому юридическую ответственность несет лицо, обязанное осуществлять человеческий контроль за использованием системы или приложения ИИ, если не выявлено иных виновных лиц. Указанный принцип постепенно воспринимается международно-правовой доктриной, что выражается в обозначении необходимости контроля за использованием ИИ в ряде международных документов.*

ВЫВОДЫ. *При условии закрепления в международном договоре в рамках ЕАЭС общеправовой принцип ответственности контролирующего лица за вред, причиненный ИИ, может приобрести значение регионального международно-правового принципа, и тем самым стать основой формирования в ЕАЭС нормативно-правового регулирования распределения ответственности за вред, причиненный ИИ. Предложенный юридический механизм пригоден для сближения законодательств государств – членов ЕАЭС как*

посредством наднационального правового регулирования, так и посредством гармонизации законодательств по данному вопросу.

КЛЮЧЕВЫЕ СЛОВА: *вред, причиненный искусственным интеллектом, разрыв ответственности, деликт, юридический состав, человеческий контроль за использованием ИИ, общеправовой принцип, принцип международного права, ЕАЭС, наднациональное правовое регулирование*

ДЛЯ ЦИТИРОВАНИЯ: Мельникова Е.Н. 2024. Принцип ответственности контролирующего лица как основа для устранения «разрыва ответственности» за вред, причиненный искусственным интеллектом. – *Московский журнал международного права.* № 4. С. 132–145. DOI: https://doi.org/10.24833/0869-0049-2024-4-132-145

Автор заявляет об отсутствии конфликта интересов.

DOI: https://doi.org/10.24833/0869-0049-2024-4-132-145

Research article
UDC: 341
Received 19 June 2024
Approved 14 November 2024

Elena N. MELNIKOVA
Saint Petersburg University
7–9, Universitetskaya Emb., Saint Petersburg, Russia, 199034
melnikova_elena5@mail.ru
ORCID ID: 0000-0001-7263-5281

PRINCIPLE OF RESPONSIBILITY OF THE CONTROLLING PERSON AS APPROACH TO ELIMINATING THE "RESPONSIBILITY GAP" FOR HARM CAUSED BY AI SYSTEMS AND AI APPLICATIONS

INTRODUCTION. *The use of artificial intelligence technologies (hereinafter referred to as "AI") is characterized by the mediation of human actions by autonomous processes, which leads, in the case when technical expertise is unable to identify the causer of harm, to a "responsibility gap" is an undesirable legal phenomenon in which the imposition of responsibility for harm caused by the use of AI on a specific person (persons) for the rules of tort liability are impossible.*

MATERIALS AND METHODS. *The research used general scientific and special methods, including the historical method, methods of formal logic, analysis, synthesis, as well as systemic and comparative legal methods.*

RESEARCH RESULTS. *To eliminate the "responsibility gap", the article proposes a mechanism that allows to fill in the missing elements of a tort committed using AI when the error that led to harm cannot be attributed de lege lata to any participant in the life cycle of an AI system or application. The starting point for the development of this mechanism was the theory of "guidance control" over the use of AI. A legal understanding of the philosophical foundations of the theory of "guidance control" allows us to substantiate the general legal principle of allocating responsibility for harm caused by AI, according to which the legal responsibility is borne by the person obliged to exercise human control over the use of the AI system or application,*

unless other perpetrators are identified. This principle is gradually being accepted by the international legal doctrine, which is expressed in the designation of the need to control the use of AI in a number of international documents.

CONCLUSIONS. *Provided that the protocol to the Treaty on the EAEU enshrines the general legal principle of responsibility of the controlling person for harm caused by AI, it can acquire the significance of a regional international legal principle, and thereby become the basis for the formation of regulatory regulation in the EAEU of the distribution of responsibility for harm caused by AI. The proposed toolkit is convenient for legal consolidation through supranational legal regulation.*

KEYWORDS: *tort, harm caused by artificial intelligence, responsibility gap, legal composition, human control over the use of AI, general legal principle, principle of international law, EAEU*

FOR CITATION: Melnikova E.N. Principle of Responsibility of the Controlling Person as Approach to Eliminating the "Responsibility Gap" for Harm Caused by AI Systems and AI Applications. – *Moscow Journal of International Law.* 2024. No. 4. P. 132–145. DOI: https://doi.org/10.24833/0869-0049-2024-4-132-145

The author declares the absence of conflict of interest.

1. Введение

Использование технологий ИИ характеризуется опосредованием действий человека автономными процессами, что приводит в ряде случаев к размыванию подотчетности действий участников жизненного цикла системы или приложения ИИ и такому нежелательному правовому явлению, при котором возложение ответственности за вред, причиненный использованием ИИ, на конкретное лицо (лиц) по правилам деликтной ответственности невозможно ввиду невозможности установления таких элементов состава гражданско-правового нарушения как причинно-следственная связь между вредом и действиями конкретных лиц, а также противоправности и виновности этих действий. Указанное правовое явление, главной характеристикой которого является затруднение поиска причинителя вреда, причиненного ИИ, принято обозначать в англоязычной литературе термином «разрыв ответственности» [Matthias 2004:175]; [De Sio, Mecacci 2021:34, 1057-1084]. «Разрыв ответственности» обусловлен такими технологическими причинами ошибок ИИ, как: 1) объективная невозможность определить причинителя вреда ввиду возникновения неблагоприятных последствий имманентной ошибки модели ИИ, действующей в составе системы ИИ, обусловленной вероятностной природой выдачи результата и поэтому не зависящей от контролирующего субъекта (истинный пробел ответственности); 2) 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 (мнимый пробел ответственности); субъективная невозможность определить причинителя вреда ввиду возникновения неблагоприятных последствий использования приложения ИИ тогда, когда работа системы ИИ, управляющей приложением ИИ, опосредуется действиями многих участников жизненного цикла приложения ИИ (размывание ответственности). «Разрыв ответственности» вызывает правовую неопределенность, провоцирующую безнаказанное причинение вреда использованием ИИ. В настоящей статье предложен юридический механизм, направленный на устранение «разрыва ответственности». Суть данного механизма заключается в трансформации моральной ответственности за выбор инструментов для принятия человеческих решений, основанных на технологиях ИИ, в юридическую ответственность.

2. Общеправовой принцип ответственности контролирующего лица за вред, причиненный использованием ИИ

22 марта 2023 г. некоммерческой организацией Future of Life опубликовано обращение, в котором глава SpaceX Илон Маск, соучредитель Apple Стив Возняк, филантроп Эндрю Янг, самый

цитируемый ученый в области IT Йошуа Бенгио и еще около тридцати тысяч исследователей ИИ призвали «немедленно приостановить» обучение систем ИИ, более мощных, чем GPT-4, предупреждая о ряде негативных последствий для человечества, среди которых упоминается даже «риск потери контроля над цивилизацией»[1].

Всеобщность и масштаб проблемы распределения ответственности за вред, причиненный ИИ, создает потребность в ориентире, снимающем правовую неопределенность в этом вопросе. Представляется, что таким ориентиром может стать общеправовой принцип ответственности контролирующего лица.

Для возникновения предпосылок к становлению общеправового принципа ответственности контролирующего лица за вред, причиненный использованием ИИ, необходимо наметить детерминанты: 1) между контролем и моральной ответственностью за выбор средств принятия решений; 2) между контролем и юридической ответственностью.

a. Взаимосвязь контроля и моральной ответственности в свете концепции значимого человеческого контроля

Взаимосвязь контроля и моральной ответственности традиционно исследуется в теории «руководящего контроля», популярной в инженерии и психологии дорожного движения начиная с 60-х годов XX в. [Michon 1985:500; Fischer, Ravizza 1998; Bovens, Miceli 1998; Collingridge 1980; Danaher 2016:299-309; Pesch 2015:925-939; Van de Poel, Sand 2018]. Теория «руководящего контроля» концентрируется на исследовании взаимосвязи моральной ответственности и контроля человека за *своими* действиями, не опосредованными автономными процессами, и не исследует взаимосвязь между контролем и юридической ответственностью.

Автономные процессы усложнили процесс установления подотчетности действий, поэтому на основе теории «руководящего контроля» юристами-международниками и исследователями этики ИИ для устранения «разрыва ответственности» за вред, причиненный ИИ, была предложена концепция значимого человеческого контроля за использованием ИИ, представляющая собой систему взглядов, в основе которых лежит идея о том, что люди должны сохранять контроль над автономными системами и нести моральную ответственность за последствия их использования [De Sio, van den Hoven 2018:11; Cavalcante, Lupetti, Aizenberg 2023:241-255; Heyns 2014; Meloni 2016; Matthias 2004:175; Sparrow 2007:62-77].

Характерно то, что в концепции значимого человеческого контроля нет согласованности в отношении ее основных элементов: определения понятия значимого человеческого контроля, условий контроля, при которых он может считаться значимым с точки зрения моральной ответственности, и даже в отношении вопроса над чем именно такой контроль может осуществляться[2]. Разнонаправленность взглядов не позволяют концепции значимого человеческого контроля развиться в научную теорию, в рамках которой может быть установлена взаимная связь человеческого контроля и моральной ответственности.

b. Взаимосвязь контроля и юридической ответственности в свете концепции значимого человеческого контроля

Взаимосвязь человеческого контроля ИИ и юридической ответственности исследуется менее активно: довольно мало исследований посвящено именно этому вопросу [Matthias 2004:175; Amoroso, Tamburrini 2020:187-194]. Вместе с тем в международном праве постепенно накапливаются разрозненные правовые нормы, которые связывают ответственность за последствия использования ИИ с фактом осуществлением контроля человеком. В частности, о необходимости человеческого контроля за использованием ИИ говорится в Резолюции Европейского парламента «О правилах гражданского

[1] Pause Giant AI Experiments: An Open Letter. March 22, 2023. URL: https://futureoflife.org/open-letter/pause-giant-ai-experiments/ (дата обращения: 28.03.2023). На 17 сентября 2024 г. под обращением подписались **33 707** человек, в том числе соучредитель Pinterest Эван Шарп, соучредитель Ripple Крис Ларсен, генеральный директор Stability AI Эмад Мостак, а также исследователи из DeepMind, Гарварда, Оксфорда и Кембриджа. Письмо подписали также тяжеловесы в области искусственного интеллекта Йошуа Бенджио и Стюарт Рассел.

[2] The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward. – *UNIDIR* : [сайт]. 13 November 2014. URL: https://unidir.org/publication/the-weaponization-of-increasingly-autonomous-technologies-considering-how-meaningful-human-control-might-move-the-discussion-forward/ (дата обращения: 23.09.2024).

права в области робототехники» от 16 февраля 2017 г. № 2015/2103(INL)[3] и Рекомендациях Совета Организации экономического сотрудничества и развития по ИИ от 2019 г.[4], принцип ответственности за контроль проводится в Регламенте Европейского союза (далее – ЕС) об ИИ от 2024[5].

Согласно ст. 9 Рамочной конвенции Совета Европы об ИИ, правах человека, демократии и верховенстве закона 2024 г.[6] государства-участники обязаны принимать меры по обеспечению подотчетности и ответственности за неблагоприятное воздействие на права человека, демократию и верховенство закона, возникающее в результате деятельности в рамках жизненного цикла систем ИИ.

В 2015 г. ряд авторов [Čerka, Grigienė, Sirbikytė 2015:376] предложили в качестве такого ориентира ст. 12 Конвенции Организации Объединенных Наций (дале – ООН) об использовании электронных сообщений в международных договорах, в которой говорится, о том, что физическое или юридическое лицо, запрограммировавшее компьютер, несет конечную ответственность за все сообщения, генерируемые машиной. Это согласуется с принципом, согласно которому владелец инструмента всегда отвечает за его использование. В данной норме четко прослеживается универсальная установка участников ООН на атрибутирование последствий ситуации контролирующему лицу.

Другая норма-ориентир по вопросу об ответственности содержится в п. 35 документа ЮНЕСКО «Рекомендации об этических аспектах ИИ»[7]: всегда должна существовать возможность возложения этической и правовой ответственности на конкретное физическое или действующее юридическое лицо.

Принципы человеческого контроля также сформулированы в документе «Азиломарские принципы»[8], принятом профессиональным IT-сообществом в 2017 г. Два из них – Human Control и Recursive self-improvement – определяют, что такое контроль ИИ человеком: «Люди должны определять процедуру и степень необходимости передачи системе ИИ функции принятия решений для выполнения целей, поставленных человеком»; «Системы ИИ, разработанные для улучшения эффективности собственных алгоритмов и самовоспроизведения, ведущего к быстрому изменению качества и количества, должны быть объектом применения мер жесткого регулирования и контроля».

Наиболее активно проблемы человеческого контроля ИИ обсуждаются в связи с применением автономного летального оружия. Именно автономное летальное оружие в 2012–2013 гг. стало поводом появления в отчетах Британской NGO Aticle 36 термина «значимый человеческий контроль». Article 36 обосновывала аргументы в пользу необходимости позитивного обязательства в международном праве, согласно которому отдельные атаки должны находиться под контролем человека.

Human Rights Watch в обзоре позиций различных государств – участников КНО по вопросу человеческого контроля смертоносных автономных систем (далее – САС) приводит полемику, которая свидетельствует о существовании четырех позиций.

---

[3] Civil Law Rules on Robotics: Resolution with rec-ommendations to the Commission N 2015/2103(INL): adopted by European Parliament 16.02.2017. – *EURO-PARL.EUROPA.EU* : [сайт]. URL: https://www.europarl.europa.eu/ doceo/document/TA-8-2017-0051_EN.html (дата обращения: 01.02.2024).

[4] Recommendation on Artificial Intelligence № OECD/LEGAL/0449: adopted by OECD Council 22.05.2019. – *LEGALINSTRUMENTS.OECD.ORG* : [сайт]. URL: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (дата обращения: 01.02.2024).

[5] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 July 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) № 300/2008, (EU) № 167/2013, (EU) № 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (OJ L, 12.7.2024, p. 1-144). – *Offic. J. of the Europ. Union.* Ser. C. L. 2024/1689. 12.07.2024. P. 1-144.

[6] Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. Council of Europe Treaty Series №. 225. Vilnius, 5.IX.2024. – *Council of Europe :* [сайт]. URL: https://rm.coe.int/1680afae3c (дата обращения: 09.09.2024).

[7] Проект рекомендации об этических аспектах искусственного интеллекта № SHS/BIO/ AHEG-AI/2020/4 REV.2: подготовлен 07.09.2020. – *UNESDOC.UNESCO.ORG* : [сайт]. URL: https://unesdoc.unesco.org/ark:/48223/pf 0000373434_rus (дата обращения: 01.02.2024).

[8] Asilomar AI Principles, coordinated by FLI and developed at the Beneficial AI 2017 conference, are one of the earliest and most influential sets of AI governance principles. 11.08.2017. URL: https://futureoflife.org/open-letter/ai-principles/ (дата обращения: 01.08.2023).

1. Наиболее радикальная: полный запрет автономных систем вооружения (например, Австрия).

2. Концепция значимого контроля человека за использованием ИИ, которой придерживается ряд государств, например: Бразилия, Болгария, Дания, Япония, Чехия, Аргентина, Мексика и др. Эти государства считают необходимым изменение норм международного права.

3. Государства, признающие необходимость человеческого контроля над САС, но не видящие необходимости изменения норм международного права. К ним относятся Российская Федерация, Китай, Швейцария, Люксембург, Иран, Болгария. Позицию этих государств наиболее последовательно выражает Российская Федерация. В ноябре 2022 г. представитель России К.В. Воронцов на совещании государств – участников Конвенции о «негуманном» оружии (далее – КНО) отметил, что Россия видит «востребованность в продолжении дискуссии о необходимости обеспечения взаимодействия «человек – машина», «сохранение контроля человека над машиной является важным условием обеспечения выполнения существующих норм международного права, в том числе МГП», «при этом конкретные формы и методы такого взаимодействия должны оставаться на усмотрение государств»[9]. Оценивая позицию сторонников концепции значимого человеческого контроля, представитель России справедливо заявил, что «продвигаемая рядом стран концепция «значимого человеческого контроля» в целом не имеет отношения к праву и чревата лишь политизацией дискуссии[10]. Позднее данные позиции были подтверждены в рабочем документе группы правительственных экспертов по новым технологиям в сфере создания САС от 14 мая 2024 г.[11]

4. Государства, которые не видят необходимости в человеческом контроле над автономными системами вооружения или «размывают» человеческий контроль: Франция, Германия, Израиль, США. Так, по мнению Израиля, соответствующее человеческое суждение уже заложено в разработку систем вооружения, в том числе на этапах проектирования, испытаний и развертывания, и, следовательно, необходимость в контроле со стороны человека является ненужной.[12] США и Израиль еще в 2015 г. ушли от использования термина «контроль» и заявляли о необходимости «обеспечить надлежащий уровень человеческого суждения о применении силы»[13]. Именно эта формулировка была в конечном счете закреплена в директиве Департамента оборонной политики по автономному оружию[14].

Особое место занимает Великобритания, первое десятилетие XXI в. разделявшая концепцию значимого человеческого контроля. Но в 2022 г. ее позиция изменилась. В документе оборонной стратегии[15] в отношении применения САС вместо понятия «значимый человеческий контроль» появилось понятие "context-appropriate

[9] Выступление К.В. Воронцова на Совещании государств-участников Конвенции о «негуманном» оружии по п. 7 повестки дня «Рассмотрение доклада Группы правительственных экспертов по развивающимся технологиям в области смертоносных автономных вооружений». 17.11.2022. – *Постоянный Представитель Российской Федерации при Отделении ООН и других МО в Женеве*. https://geneva.mid.ru/centr-verh/-/asset_publisher/vxGb9AKjOm4T/content/vystuplenie-k-v-voroncova-na-sovesanii-gosudarstv-ucastnikov-konvencii-o-negumannom-oruzii-po-p-7-povestki-dna-rassmotrenie-doklada-gruppy-pravitel-st (дата обращения: 20.09.2024).

[10] Там же.

[11] Рабочий документ группы правительственных экспертов по новым технологиям в сфере создания САС от 14 мая 2024 г. Совещание 4–8 марта и 26–30 августа 2024 г. по Конвенции о «негуманном» оружии. – *UNODA* : [сайт]. 14 мая 2023. URL: https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2024)/CCW-GGE.1-2024-WP.2_Russian.pdf (дата обращения: 20.09.2024).

[12] Statement by Israel. CCW Meeting of States Parties. Geneva. November 13-14, 2014. – *UNODA* : [сайт]. URL: https://unoda-documents-library.s3.amazonaws.com/Convention_on_Certain_Conventional_Weapons_-_Meeting_of_High_Contracting_Parties_(2014)/Israel_MSP_GS.pdf (дата обращения: 20.09.2024).

[13] Statement of the United States, CCW Meeting of Experts on Lethal Autonomous Weapons Systems, April 13, 2015. URL: http://www.unog.ch/80256EDD006B8954/(httpAssets)/8B33A1CDBE80EC60C1257E2800275E56/$file/2015_LAWS_MX_USA+bis.pdf (дата обращения: 02.10.2023).

[14] DOD Directive 3000.09 «Автономность в системах вооружения». January 25, 2023. – *Executive Services Directorate* : [сайт]. URL: https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf (дата обращения: 05.06.2024).

[15] Policy paper Defence Artificial Intelligence Strategy. – *GOV.UK* : [сайт]. URL: https://www.gov.uk/government/publications/defence-artificial-intelligence-strategy (дата обращения: 05.06.2024).

human involvement", которое в переводе на русский язык может быть понято как «соответствующее контексту участие человека». В указанном документе утверждается, что механизмы человеческой власти, ответственности и подотчетности всегда будут присутствовать при применении силы, однако нет указаний на то, как следует оценивать или понимать «соответствующее контексту участие человека»[16].

Государства, считающие недостаточными существующие международные нормы для контроля над САС, добиваются подписания нового протокола к КНО. 7 марта 2023 г. Управление верховного комиссара ООН по делам беженцев провело вебинар, посвященный понятию «значимого человеческого контроля» в отношении новых технологий в области САС, где «значимый контроль со стороны человека» был назван «рамочной концепцией для дискуссии об автономии систем вооружения» и центральным вопросом, «предъявляющим требования к обеспечению соответствия систем автономного оружия международному гуманитарному праву»[17]. Участники вебинара обосновывали условия, при которых человеческий контроль может считаться значимым, причем доминировали разнообразные условия технического взаимодействия «человек – машина».

Проведенный анализ позволяет прийти к выводу о том, что в рамках ***концепции значимого человеческого контроля за использованием ИИ предпринимаются упорные попытки юридической квалификации факта взаимодействия «человек – машина» в соответствии с нормами международного права,*** однако это осложняется тем, что «у политиков и проектировщиков отсутствует разработанная теория о том, что именно означает «значимый контроль» со стороны человека» [De Sio, van den Hoven 2018:3] и поэтому они не знают, «какие конкретные правовые нормы и правила должны вытекать из этого принципа» [De Sio, van den Hoven 2018:3]. Приходится констатировать, что концепция значимого контроля человека за использованием ИИ не разработана, а ее главный постулат – отвечает тот, кто должен контролировать, – пока не получил юридического обоснования, достаточного для построения взаимосвязи между контролем и юридической ответственностью.

С учетом состояния мирового правопорядка следует согласиться с тем, что на сегодняшний день являются «преждевременными дискуссии по нормам, принципам и правилам «ответственного поведения» в отношении САС»[18]. Однако правовое регулирование ИИ гражданского назначения является не менее важным, поскольку появление систем и приложений ИИ – явление, имеющее значение для всего международного сообщества ввиду своей способности к трансграничному причинению вреда [Мельникова 2024:95] и неконтролируемой трансграничной пролиферации [Burri 2017:97], способной затронуть различные сферы жизни гражданского общества.

с. Понятие человеческого контроля за использованием ИИ и условия его осуществления

Для устранения «разрыва ответственности» необходимо разработать юридический подход, который одновременно соответствовал бы философским основаниям ответственности, был бы физически возможным и при этом непротиворечивым с точки зрения права. Для этого необходимо определить понятие и условия значимого человеческого контроля.

Немецким исследователем в области этики ИИ С. Робинсоном был предложен подход к выявлению смысла термина «значимый человеческий контроль». Обобщив наработки компатибилистской теории «руководящего контроля» Фишера и Равиццы [Fischer, Ravizza 1998], ученый сформулировал три условия наступления моральной ответственности за вред, причиненный ИИ [Robbins 2023], на основе которых можно сформулировать три условия возможности осуществления человеческого контроля.

---

[16] Defence Artificial Intelligence Strategy. – *Article 36* : [сайт]. June 2022. URL:https://article36.org/updates/new-uk-government-position-on-autonomous-weapons-recognises-that-lines-need-to-be-drawn-but-lacks-detail-or-signs-of-real-leadership/ (дата обращения: 02.01.2024).

[17] Neder C. Experts Reflect on Novel Approaches to «Meaningful Human Control». – *UNODA* : [сайт]. 7 March 2023. URL: https://disarmament.unoda.org/the-convention-on-certain-conventional-weapons/experts-reflect-on-novel-approaches-to-meaningful-human-control/(accessed date: 23.09.2024).

[18] Рабочий документ группы правительственных экспертов по новым технологиям в сфере создания САС от 14 мая 2024 г.

Условие 1: Осознанность. *Контроль должен быть осознанным, т. е. осуществляться специальным субъектом – лицом, обладающим квалификацией достаточной для предвидения последствий действий своих и системы ИИ и способным к осуществлению действий по изменению результатов вывода системы или приложения ИИ. Представляется, что контроль не должен рассматриваться через призму его значимости: факт контроля либо есть, либо нет, контроль не может быть двух видов – значимый и не значимый. На сегодняшний день условия человеческого контроля формулируемые через категорию «значимости» каждое государство понимает так, как ему выгодно. Вместо этого целесообразно пользоваться критерием «осознанности», достаточной для предвиденья последствий действий своих и системы ИИ.*

Условие 2: «Человекоразмерность». *Для соблюдения условия 1 система ИИ или приложение ИИ должны быть человекоразмерными, т. е. позволяющими изменить свои результаты с учетом психофизиологической скорости принятия решения человеком.*

Условие 3. Прослеживаемость. *Возможность «прослеживаемости» результата операций системы ИИ для установления технологических причин ошибки и определения вида «разрыва ответственности» с целью его устранения возложением ответственности на лицо, осуществляющее человеческий контроль.*

Условие 1 является главным и необходимым. Условия 2 и 3 должны определять законные пределы делегирования человеческих полномочий машине (что является темой для отдельного исследования).

Принимая во внимание вышеперечисленные условия, ***человеческий контроль за использованием ИИ может быть определен как осуществление лицом действий, способных влиять на вывод результата системой ИИ, управляющей приложением ИИ с учетом психофизиологической скорости принятия решения человеком, обладающим квалификацией, достаточной для предвиденья последствий действий своих и системы (приложения) ИИ; результаты выводов системы ИИ должны обладать свойством прослеживаемости.***

Человеческий контроль в «узком» значении. Если в отношении деятельности лица выполняются все вышеуказанные условия, то такая деятельность может рассматриваться в качестве значимого человеческого контроля за использованием ИИ в «узком» значении. В этом случае «разрыв подотчетности» отсутствует[19], и, соответственно, «разрыв ответственности», как моральной, так юридической, не образуется.

Человеческий контроль в «широком» значении. Если не выполняется условие 1, то это признак либо наличия имманентной ошибки модели ИИ, либо нарушения приделов делегирования человеческих полномочий машине (в последнем случае также не соблюдаются условия 2 и 3). В этом случае имеется «разрыв подотчетности», следствием которой являются затруднения в атрибутировании ошибки конкретному лицу, поскольку условия осуществления контроля в «узком» значении не выполняются. Для устранения «разрыва ответственности» в случае невыполнения условий значимого человеческого контроля в «узком» значении важно обозначить понятие значимого человеческого контроля за использованием ИИ в его *широком значении,* когда контролирующее лицо, по сути, не может повлиять на негативный результат использования ИИ иначе, кроме как отказом от такого использования (так как технически устранить «разрыв ответственности» невозможно). Здесь проблема ответственности вызвана тем, что контролирующее лицо должно отвечать (сначала морально, а потом уже юридически) за чужие действия, за чужие ошибки. Вопрос о моральной ответственности в таком случае является предметом текущих философских дебатов [De Sio, van den Hoven 2018:4]. Думается, что наиболее глубоко обусловленность моральной ответственности возможностью осуществления контроля как *выбора механизма принятия решений* обоснована в компатибилистской теории «руководящего контроля» Фишера и Равиццы [Fischer, Ravizza 1998]. Указанная теория руководящего контроля описывает условия, при которых агенты-люди несут моральную ответственность за свои действия, основываясь на особенностях механизма принятия решений, приводящего к этим действиям.

[19] Имманентная ошибка модели, и проблема «черного ящика» для рассматриваемой ситуации не характерна, затруднения в атрибутировании ошибки отсутствуют.

Первое условие Фишера и Равиццы для руководящего контроля – условие разумной реакции [Fischer, Ravizza 1998:6]. Оно требует, чтобы механизм принятия решений агентом был чувствителен и реагировал на достаточное разнообразие моральных факторов, т. е. чтобы механизм принятия решений мог адаптировать поведение системы ИИ к моральным особенностям обстоятельств. Однако на данном этапе развития уровня техники это условие пока не выполнимо [Melnikova, Surov 2023:22].

Второе условие Фишера и Равиццы основывается на доктрине причинного детерминизма и применительно к руководящему контролю состоит в том, что нести моральную ответственность за свои действия может только тот агент, который несет в себе механизм принятия решений: агент должен «взять на себя ответственность» за тот механизм, с помощью которого он принимает решения [Fischer 2004:146]. Это означает, что субъект может и должен предвидеть, как тот или иной механизм принятия решений, который он выбрал, способен повлиять на мир. Если субъект опосредует принятие своих решений технологиями ИИ, то за этот выбор он должен нести моральную ответственность.

Исходя из вышеизложенного, *человеческий контроль в «широком» значении* может быть определен как такая деятельность профессионального пользователя ИИ, при которой не соблюдаются одно или несколько условий человеческого контроля (осознанность, человекоразмерность, прослеживаемость)[20], однако, зная о вероятностном характере вывода результата моделями ИИ, лицо тем не менее опосредует механизм принятия решений использованием технологий ИИ.

Философские выводы теории «руководящего контроля», обоснование понятия контроля в «широком» значении, в совокупности с положениями международно-правовых актов, указывающих на необходимость контроля за ИИ, позволяют прийти к выводу о существовании жесткой детерминанты между контролем и юридической ответственностью.

Необходимость ответственности за выбор средств принятия человеческих решений, опосредованных технологиями ИИ, может быть сформулирована в виде ***общеправового принципа распределения ответственности за вред, причиненный ИИ, согласно которому юридическую ответственность несет лицо, обязанное осуществлять человеческий контроль за использованием системы или приложения ИИ, если не выявлено иных виновных лиц.***

Вместе с тем обоснование вышеуказанного принципа само по себе не решает проблему «разрыва ответственности», а лишь является первым шагом к ее решению. Создание юридического механизма устранения «разрыва ответственности» за вред, причиненный ИИ, требует выполнения дополнительных шагов к трансформации моральной ответственности за использование ИИ в юридическую. В настоящей статье рассмотрены вопросы только гражданско-правовой ответственности.

3. Фикция осуществления значимого человеческого контроля в широком значении

Вторым шагом является нормативное закрепление обязанности по осуществлению человеческого *контроля* в «широком» значении за лицом, опосредующим принятие решений технологиями ИИ. Это может быть достигнуто введением в правовое регулирование *фикции* осуществления человеческого контроля за использованием ИИ в широком значении, на основе которой можно сформировать презумпции, необходимые для восполнения тех элементов состава гражданско-правового деликта, которые иначе невосполнимы в случае «разрыва ответственности».

Классическим в юридической науке считается объяснение, что презумпция не может строиться на основе фикции, а может строиться только на основе существующего факта. Проблема в том, что *de lege lata* в случае деликта с участием ИИ не всегда возможно установление факта, на основе которого можно построить презумпции, восполняющие элементы состава.

Полагая, что в общественных науках, особенно в юриспруденции, фикции не должны быть противоестественными, поскольку их применение может влиять на основные права человека и благополучие окружающей среды, мы придерживаемся «непопулярной» позиции, согласно которой «презумпция есть предположение о наличие факта, вероятность существования

[20] Соответственно, контроль в его «узком» значении невозможен.

которого велика. Фикция является предположением о несуществующем факте или о факте, вероятность которого мала или неизвестна» [Черниловский 1984].

Таким фактом мы предлагаем считать осуществление человеческого контроля в широком значении, которое ввиду относительной новизны и отсутствия общепринятого целенаправленного и четкого философского обоснования ряда вышеупомянутых детерминант на сегодняшний день может принимать только форму фикции.

Думается, что второе условие руководящего контроля Фишера и Равиццы [Fischer, Ravizza 1998:19], согласно которому человек должен «взять на себя ответственность» за то средство, с помощью которого он принимает решения тогда, когда *проявлена неосторожность в выборе средства,* применительно к ИИ может послужить основой для формирования *фикции осуществления человеческого контроля в широком значении. Широкий подход к пониманию человеческого контроля через ответственность человека за выбор средств принятия решений* позволяет восполнить недостаток возможности предвидения последствий использования системы ИИ, недостаток «человекоразмерности» системы или приложения ИИ (в случае нарушения пределов делегирования человеческих полномочий машине), недостаток прослеживаемости ошибок.

Применение фикции осуществления человеческого контроля в широком значении целесообразно там, где не выполняются условия человеческого контроля в «узком» значении.

Функция человеческого контроля в широком значении позволяет привлечь к гражданско-правовой ответственности лицо тогда, когда контроль в «узком» значении был невозможен[21]. Лицо может рассматриваться как допустившее *неосторожность* в использовании системы или приложения ИИ *на этапе выбора* средства достижения поставленных целей.

Фикция осуществления человеческого контроля в широком значении не является противоестественной, непротиворечива с точки зрения теории права: через какое-то время не исключен переход осуществления человеческого контроля в широком значении в разряд юридических фактов. Поэтому представляется возможным построить на основе фикции осуществления человеческого контроля в широком значении необходимые презумпции для восполнения недостающих элементов состава деликта.

4. Презумпции, восполняющие недостающие элементы юридического состава деликта, опосредованного использованием ИИ, когда экспертиза не способна атрибутировать ошибку конкретному лицу

Третьим завершающим шагом к устранению «разрыва ответственности» является установление следующих презумпций.

Презумпция вины в форме неосторожности контролирующего лица в выборе средства принятия решения, опосредованного технологиями ИИ

В случае вреда, причиненного использование ИИ, вина лица, осуществляющего человеческий контроль, может иметь место, но ее может и не быть.

Представляется, что виновное поведение контролирующего лица соответствует случаю осуществления контроля в «узком» значении, поскольку: а) гражданское законодательство предусматривает презумпцию вины причинителя вреда; б) ввиду того, что в гражданском праве под виной причинителя вреда понимают «непринятие им надлежащих мер по устранению или недопущению отрицательных результатов своих действий, диктуемых обстоятельствами конкретной ситуации» [Гражданское право 2000:449].

Сложнее обстоят дела с установлением виновного поведения лица в случае осуществления им человеческого контроля за использованием ИИ в «широком» значении, поскольку в этом случае контролирующее лицо отвечает за чужие действия (так как имеет место имманентная ошибка модели ИИ, или контроль осуществляет лицо без соответствующей квалификации, или система или приложение ИИ в принципе

[21] Соответственно, в отношении работников организаций, которые не могут осуществлять выбор – использовать ИИ или нет, целесообразно говорить о надзоре за использованием ИИ. По этой же причине – невозможности осуществления выбора, – нельзя говорить о человеческом контроле за использованием ИИ и со стороны потребителей, которые приобретают то, что предлагает рынок. Контроль в данном случае должно осуществлять лицо, которое вывело такие продукты на рынок.

не поддается контролю по причине ее «нечеловекоразмерности», либо проследить причину ошибки невозможно). Думается, в этом случае может презюмироваться вина в форме неосторожности, допущенной контролирующем лицом, когда он выбрал технологию ИИ в качестве средства, опосредующего принятие решений.

Соответственно, вина контролирующего лица должна презюмироваться всегда, независимо от того, «узкий» или «широкий» подход к понятию контроля мы применяем.

При осуществлении контроля в «широком» значении, в том числе, когда экспертизой установлена имманентная ошибка модели ИИ (чему соответствует «истинный пробел ответственности»), презумпция должна быть *неопровержимой*, поскольку лицо, использующее ИИ, должно быть осведомлено о том, что оно должно будет принять риск вероятностной ошибки модели ИИ, повлекший неправильный вывод результата системой ИИ.

Применение юридической фикции осуществления человеческого контроля *в широком значении* и *неопровержимой* презумпции неосторожности должно осуществляться *в исключительном случае,* когда существует *объективная* невозможность атрибутирования ошибки человеку, при этом нужно устранить «истинный пробел ответственности».

Неоднозначная ситуация возникает в случае субъективной невозможности атрибутирования ошибки, когда экспертиза не может достоверно установить либо причину ошибки (проблема «черного ящика» и «мнимый пробел ответственности»), либо лицо, действиями которого вызвана ошибка (проблема «многих рук» и «размывание ответственности»). В этом случае лицо, причинившее вред, объективно существует, просто из-за технической сложности его невозможно определить. Поскольку атрибутировать ошибку в обоих случаях практически невозможно, то и правовые последствия целесообразно сделать сходными с ситуацией, когда имеет место установленная вероятностная ошибка модели ИИ. Иными словами, для устранения как мнимого «разрыва ответственности», так «размывания ответственности» следует применить фикцию осуществления человеческого контроля *в широком значении,* но при этом *опровержимую* презумпцию неосторожности пользователя (эксплуатанта) системы или приложения ИИ.

Презумпция наличия причинно-следственной связи между вредом и действиями лица, обязанного осуществлять контроль, и противоправности этих действий

Осуществление человеческого контроля в «широком» значении предполагает неосторожность в выборе средства принятия решения опосредованием технологий ИИ, что позволяет презюмировать противоправность неосторожных действий даже в случае соблюдения всех обязанностей и инструкций. Презюмируя неосторожность действий, мы презюмируем и наличие причинно-следственной связи между неосторожным поведением контролирующего лица, ошибкой и наступившими последствиями (вредом).

Чтобы освободиться от ответственности владельцу системы или приложения ИИ, посредством которых причинен вред, необходимо доказать, что вред причинен другим лицом, а именно: а) лицо, которому вменяется осуществление человеческого контроля в «широком» значении может доказать, что существует лицо, осуществляющее человеческий контроль в «узком» значении; б) лицо, которому вменяется осуществление человеческого контроля (в широком или узком значении), может доказать, что существует иное лицо, причинившее вред. Соответственно, предлагаемый подход не противоречит гражданско-правовому постулату: лицо, причинившее вред, освобождается от обязанности его возмещения, если докажет, что вред причинен не по его вине.

Поскольку фикция человеческого контроля в «широком» значении неприменима к физическим лицам, неприменимы к ним и вышеуказанные презумпции.

5. Заключение

Человеческий контроль за использованием ИИ может быть определен как осуществление лицом действий, способных влиять на вывод результата системой ИИ, управляющей приложением ИИ с учетом психофизиологической скорости принятия решения человеком, обладающим квалификацией, достаточной для предвиденья последствий действий своих и системы (приложения) ИИ; результаты выводов системы ИИ должны обладать свойством прослеживаемости.

Юридическое осмысление философских оснований теории «руководящего контроля» позволяет обосновать **общеправовой принцип распределения ответственности** за вред, причиненный ИИ, согласно которому юридическую

ответственность несет лицо, обязанное **осуществлять человеческий контроль** за использованием системы или приложения ИИ, если не выявлено иных виновных лиц. Указанный принцип постепенно воспринимается международно-правовой доктриной, что выражается в обозначении необходимости контроля за использованием ИИ в ряде международных документов. Предложенный в настоящей работе механизм устранения «разрыва ответственности» концентрирует ответственность на лице, осуществляющем человеческий контроль за использованием ИИ. Указанное лицо может освободиться от ответственности, если докажет, что вред причинило другое лицо. Общеправовой принцип ответственности контролирующего лица задает направление правового регулирования распределения ответственности за вред, причиненный ИИ, что достигается в том числе посредством установления позитивных прав и обязанностей участников жизненного цикла систем и приложений ИИ.

Для реализации указанного принципа необходим механизм возложения ответственности на контролирующее лицо исходя из имеющихся или потенциальных юридических инструментов соответствующей отрасли права. В статье предложены возможные гражданско-правовые презумпции, однако в других отраслях права сконструировать механизм возложения ответственности на контролирующее лицо может оказаться проблематично.

Именно сейчас, когда правовое регулирование ИИ только формируется, очень важно в ЕАЭС не ошибиться с правилами распределения ответственности за вред, причиненный ИИ. В условиях единого внутреннего рынка ЕАЭС системы и приложения ИИ свободно распространяются, поэтому «разрыв ответственности» даже в одном из государств-членов способен негативно сказаться и на интересах тех государств-членов, где правовое регулирование его исключает. В настоящее время в ряде стран ЕАЭС специальные нормы, посвященные правовому регулированию ИИ (действующие и планируемые), имеют противоположную направленность. Сравнительный анализ законодательного регулирования ИИ в государствах ЕАЭС показал сильную дифференциацию подходов, в частности, по вопросам допуска иностранного капитала на рынок информационных технологий и государственно-административного регулирования ИИ. Неоднородность правового регулирования ответственности за вред, причиненный ИИ, может привести к тому, что целые страны с благоприятными юрисдикциями[22] могут стать площадками для экспериментов и разработки небезопасного ИИ, в том числе иностранного происхождения. При таких обстоятельствах неоднородность правового регулирования распределения ответственности за вред, причиненный ИИ, в условиях таможенного союза способна ослабить «цифровой суверенитет» всех государств – членов ЕАЭС.

Поэтому в ЕАЭС желательно и целесообразно достижение единого юридического механизма устранения «разрыва ответственности» за вред, причиненный ИИ. Это является необходимым условием формирования нормативно-правового регулирования ИИ.

Наднациональное правовое регулирование ИИ, в основу которого положен общеправовой принцип ответственности контролирующего лица, позволит избежать дифференциации подходов к распределению ответственности за вред, причиненный ИИ. Путь от формирования общеправового принципа до его признания в качестве принципа международного права может быть долгим. При условии закрепления в международном договоре в рамках ЕАЭС общеправовой принцип ответственности контролирующего лица за вред, причиненный ИИ, приобретет значение регионального международно-правового принципа. Однако есть серьезное препятствие: на сегодняшний день отсутствуют предпосылки формирования воль государств – членов ЕАЭС на сближение законодательства стран ЕАЭС в части правового регулирования ИИ, что позволяет констатировать отсутствие предпосылок единой, согласованной и скоординированной политики стран ЕАЭС в вопросах создания и использования ИИ, а следовательно, и перспектив сближения законодательства в рассматриваемой сфере общественных отношений посредством формирования наднационального правового регулирования. Вместе с тем право ЕАЭС не содержит к этому и правовых препятствий [Мельникова 2024:98]. Поэтому в случае

[22] Где поощряется безнаказанное использование ИИ, когда причинителей вреда либо не установить вовсе, либо ответственность распределяется между широким кругом субъектов.

проявления политической воли на сближение законодательств государств – членов ЕАЭС по данному вопросу, предложенный в статье механизм пригоден для реализации его в праве ЕАЭС посредством инструментов унификации и гармонизации законодательств как в рамках единой, так и согласованной политики.

Список литературы

1. Гражданское право. Учебник. 2000. В 2-х т. Т. 1. Под ред. Е.А. Суханова. 2-е изд. Москва: БЕК. 704 с.
2. Мельникова Е.Н. 2024. Перспективы сближения законодательства стран ЕАЭС в части правового регулирования искусственного интеллекта. – *Евразийская интеграция: экономика, право, политика.* Т. 18. № 2(48). С. 95-103.
3. Черниловский З.М. 1984. Презумпции и фикции в истории права. – *Советское государство и право*. № 1. Справочная правовая система «Консультант Плюс».
4. Amoroso D., Tamburrini G. 2020. Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues. – *Current Robotics Reports.* Vol. 1. P. 187-194. DOI: 10.1007/s43154-020-00024-3.
5. Burri T. 2017. International Law and Artificial Intelligence. – *German Yearbook of International Law.* Vol. 60. P. 91-108.
6. Bovens M., Miceli M.P. 1998. The quest for responsibility: Accountability and citizenship in complex organisations. – *Cambridge University Press.* DOI: 10.2307/2667065. URL: https://www.researchgate.net/publication/275840305_The_Quest_for_Responsibility_Accountability_and_Citizenship_in_Complex_Organisations (accessed date: 02.12.2023).
7. Cavalcante S.L., Lupetti M.L., Aizenberg E. 2023. Meaningful human control: actionable properties for AI system development. – *AI Ethics.* Vol. 3. P. 241-255. DOI: 10.1007/s43681-022-00167-3.
8. Čerka P., Grigienė J., Sirbikytė G. 2015. Liability for damages caused by artificial intelligence. – *Computer Law & Security Review.* Vol. 31. Issue 3. P. 376-389.
9. Collingridge D. 1980. *The Social Control of Technology.* London: Frances Pinter. 200 p.
10. Danaher J. 2016. Robots, law and the retribution gap. – *Ethics and Information Technology.* Vol. 18. P. 299-309. DOI: 10. 1007/ s10676- 016- 9403-3.
11. De Sio S.F., Mecacci G. 2021. Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them. – *Philos and Technol.* Vol. 34. P. 1057-1084. DOI: 10.1007/s13347-021-00450-x. URL: https://link.springer.com/article/10.1007/s13347-021-00450-x (accessed date: 02.10.2023).
12. De Sio F.S., van den Hoven J. 2018. Meaningful Human Control over Autonomous Systems: A Philosophical Account. – *Front. Robot.* AI 5:15. 28 Feb. DOI: 10.3389/frobt.2018.00015.
13. Fischer J., Ravizza M. 1998. Responsibility and Control: A Theory of Moral Responsibility. – *Cambridge University Press.* 277 p. DOI:10.1017/CBO9780511814594.
14. Heyns C. Report of the Special Rapporteur on Extra-Judicial, Summary or Arbitrary Executions. – *Geneva: United Nations,* 1 Apr. 2014. URL: https://digitallibrary.un.org/record/771922?ln=ru (дата обращения: 30.12.2023).
15. Melnikova E., Surov I. 2023. Legal Status of Artificial Intelligence from Quantum-Theoretic Perspective. – *BRICS Law Journal.* Vol. X. Issue 4. P. 5-34.
16. Matthias A. 2004. The responsibility gap: Ascribing responsibility for the actions of learning automata. – *Ethics and Information Technology.* Vol. 6(3). P. 175-183. DOI: 10.1007/s10676-004-3422-1.
17. Meloni C. State and individual responsibility for targeted killings by drones. – *Drones and Responsibility: Legal, Philosophical and Socio-Technical Perspectives on Re-motely Controlled Weapons.* E. Di Nucci & De Sio F.S. (Eds.). Routledge. 2016. URL: https://www.academia.edu/77047662/State_and_individual_responsibility_for_targeted_killings_by_drones (дата обращения: 30.12.2023).
18. Michon J. A. 1985. A critical view of driver behavior models: what do we know, what should we do?– *Human behavior and traffic safety.* Ed. by Evans L., Schwing R.C. General Motors Research Laboratories. New York: Plenum Press. P. 485-524.
19. Pesch U. 2015. Engineers and Active Responsibility. – *Science and Engineering Ethics.* Vol. 21(4). P. 925-939. DOI: 10. 1007/ s11948- 014- 9571-7.
20. Robbins S. 2023. The many meanings of meaningful human control. – *AI Ethics.* DOI.10.1007/s43681-023-00320-6.
21. Sparrow R. 2007. Killer robots. – *Journal of Applied Philosophy.* Vol. 24(1). P. 62-77. DOI: 10.1111/j.1468-5930.00346.x.
22. Van de Poel I., Sand M. 2018. Varieties of responsibility: two problems of responsible innovation. – *Synthese*. DOI:10.1007/s11229-018-01951-7.

References

1. Amoroso D., Tamburrini G. 2020. Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues. – *Current Robotics Reports.* Vol. 1. P. 187-194. DOI: 10.1007/s43154-020-00024-3.
2. Bovens M., Miceli M.P. 1998. The quest for responsibility: Accountability and citizenship in complex organisations. – *Cambridge University Press.* DOI: 10.2307/2667065. URL: https://www.researchgate.net/publication/275840305_The_Quest_for_Responsibility_Accountability_and_Citizenship_in_Complex_Organisations (accessed date: 02.12.2023).
3. Burri T. 2017. International Law and Artificial Intelligence. – *German Yearbook of International Law.* Vol. 60. P. 91-108.
4. Cavalcante S.L., Lupetti M.L., Aizenberg E. 2023. Meaningful human control: actionable properties for AI system development. – *AI Ethics.* Vol. 3. P. 241-255. DOI: 10.1007/s43681-022-00167-3.
5. Čerka P., Grigienė J., Sirbikytė G. 2015. Liability for damages caused by artificial intelligence. – *Computer Law & Security Review.* Vol. 31. Issue 3. P. 376-389.
6. Chernilovsky Z.M. Prezumpcii i fikcii v istorii prava [Presumptions and fictions in the history of law]. – *Sovetskoe*

gosudarstvo i pravo [The Soviet State and law]. 1984. № 1. SPS Consultant Plus. (In Russ.).
7. Collingridge D. 1980. *The Social Control of Technology*. London: Frances Pinter. 200 p.
8. Danaher J. 2016. Robots, law and the retribution gap. – *Ethics and Information Technology*. Vol. 18. P. 299-309. DOI: 10. 1007/ s10676- 016- 9403-3.
9. De Sio S.F., Mecacci G. 2021. Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them. – *Philos and Technol*. Vol. 34. P. 1057-1084. DOI: 10.1007/s13347-021-00450-x. URL: https://link.springer.com/article/10.1007/s13347-021-00450-x (accessed date: 02.10.2023).
10. De Sio F.S., van den Hoven J. 2018. Meaningful Human Control over Autonomous Systems: A Philosophical Account. – *Front. Robot*. AI 5:15. 28 Feb. DOI: 10.3389/frobt.2018.00015.
11. Fischer J., and Ravizza M. 1998. Responsibility and Control: A Theory of Moral Responsibility. – *Cambridge University Press*. 277 p. DOI:10.1017/CBO9780511814594.
12. *Grazhdanskoe parvo. Uchebnik. V 2-h t. T. 1* [Civil law. Textbook. In 2 vol. Vol. 1]. Pod red. E.A. Suhanova. 2-e izd. Moscow: BEK. 2000. 704 p. (In Russ.).
13. Matthias A. 2004. The responsibility gap: Ascribing responsibility for the actions of learning automata. – *Ethics and Information Technology*. Vol. 6(3). P. 175-183. DOI: 10.1007/s10676-004-3422-1.
14. Melnikova E., Surov I. 2023. Legal Status of Artificial Intelligence from Quantum-Theoretic Perspective. – *BRICS Law Journal*. Vol. X. Issue 4. P. 5-34.
15. Melnikova E.N. 2024. Perspektivy sblizheniya zakonodatel'stva stran EAES v chasti pravovogo regulirovaniya iskusstvennogo intellekta [Prospects for the convergence of the legislation of the EAEU countries in terms of the legal regulation of artificial intelligence]. – E*vrazijskaja integracija: jekonomika, pravo, politika [Eurasian integration: economics, law, politics]*. Vol. 18. No. 2(48). P. 95-103. (In Russ.)
16. Meloni C. State and individual responsibility for targeted killings by drones. – *Drones and Responsibility: Legal, Philosophical and Socio-Technical Perspectives on Re-motely Controlled Weapon*. E. Di Nucci & De Sio F.S. (Eds.). Routledge. 2016. URL: https://www.academia.edu/77047662/State_and_individual_responsibility_for_targeted_killings_by_drones (дата обращения: 30.12.2023).
17. Michon J. A. 1985. A critical view of driver behavior models: what do we know, what should we do? – *Human behavior and traffic safety*. Ed. by Evans L., Schwing R.C. General Motors Research Laboratories. New York: Plenum Press. P. 485-524.
18. Heyns C. Report of the Special Rapporteur on Extra-Judicial, Summary or Arbitrary Executions. – *Geneva: United Nations,* 1 Apr. 2014. URL: https://digitallibrary.un.org/record/771922?ln=ru (дата обращения: 30.12.2023).
19. Pesch U. 2015. Engineers and Active Responsibility. – *Science and Engineering Ethics*. Vol. 21(4). P. 925-939. DOI: 10. 1007/ s11948- 014- 9571-7.
20. Robbins S. 2023. The many meanings of meaningful human control. – *AI Ethics*. DOI.10.1007/s43681-023-00320-6.
21. Sparrow R. 2007. Killer robots. – *Journal of Applied Philosophy*. Vol. 24(1). P. 62-77. DOI: 10.1111/j.1468-5930.00346.x.
22. Van de Poel I., Sand M. 2018. Varieties of responsibility: two problems of responsible innovation. – *Synthese*. DOI:10.1007/s11229-018-01951-7.

Информация об авторе

Елена Николаевна МЕЛЬНИКОВА,
соискатель ученой степени кандидата юридических наук, Санкт-Петербургский государственный университет

Университетская наб., д. 7–9, Санкт-Петербург, 199034, Российская Федерация

melnikova_elena5@mail.ru
ORCID: 0000-0001-7263-5281

About the Author

Elena N. MELNIKOVA,
Post-Graduate student, Saint Petersburg University

7–9, Universitetskaya Emb., Saint Petersburg, Russia, 199034

melnikova_elena5@mail.ru
ORCID: 0000-0001-7263-5281