НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ «ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Доклад НИУ ВШЭ

«ПРАВОВЫЕ АСПЕКТЫ ИСПОЛЬЗОВАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: АКТУАЛЬНЫЕ ПРОБЛЕМЫ И ВОЗМОЖНЫЕ РЕШЕНИЯ»

Руководители авторского коллектива: **В. Б. Наумов, С. А. Чеховская, А. Ю. Брагинец, А. В. Майоров**

Издательский дом
Высшей школы экономики
Москва, 2021

Введение

1. В рамках исследований комплексного правового регулирования перспективных направлений цифровой экономики Институтом права цифровой среды НИУ ВШЭ в 2020 году проведены научно-исследовательские работы по различным правовым аспектам использования искусственного интеллекта (ИИ). В настоящем докладе изложены результаты работ по следующим темам:

- использование ИИ при *принятии юридически значимых решений*,
- регулирование *ответственности* в сфере ИИ,
- совершение *гражданско-правовых сделок* с использованием систем ИИ,
- *алгоритмическая прозрачность* в процессах принятия решений,
- устранение барьеров для инвестиционных проектов в сфере ИИ и робототехники.

2. Именно в сфере использования ИИ верно утверждение, что право осуществляет свои регулятивные функции не изолированно и обособленно, а в едином комплексе и тесном взаимодействии с другими социальными регуляторами[1]. В этой связи требует особой проработки отражение в проектируемых нормах *этических принципов* использования систем ИИ, которые, как подтверждает мировой опыт, приобретают особое значение в этой сфере.

Использование искусственного интеллекта при принятии юридически значимых решений

3. Быстрое развитие технологий искусственного интеллекта ставит вопрос о правовом статусе использования систем ИИ, обладающих разной степенью автономности: по некоторым прогнозам, к 2075 году мыслительные процессы роботов уже нельзя будет отличить от процессов мышления человека[2]. Можно полагать, что системы ИИ могут выступать *не только как объект регулирования,* но и, зачастую одновременно, как *потенциальный инструмент применения и/или обеспечения соблюдения регулирования*[3], в частности *при принятии юридически значимых решений*, и в самое ближайшее время во всем мире будет наблюдаться процесс постепенного правового признания тех или иных «действий» системы ИИ и их последствий, а также формализации этих действий.

4. Принятие юридически значимых решений может обладать характером властного воздействия, то есть, по сути, система ИИ фактически будет *наделена властью*, в этом случае должен быть определенно решен вопрос о порядке ответственности в случае такого

[1]Нерсесянц В.С. Философия права: Учебник для вузов. М., 2004, с. 77.
[2] Etzioni Oren. No, the Experts Don't Think Superintelligent AI is a Threat to Humanity// https://www.technologyreview.com/s/602410/no-the-experts-dont-think-superintelligent-aiis-a-threat-to-humanity
[3] Regulating Artificial Intelligence// Edit. by T. Wischmeyer, T. Rademacher. Springer, 2020, p.vii// электронные ресурсы ВШЭ.

автоматизированного принятия решения. При этом юридически значимые решения – это группа решений, которая охватывает разные по своей сложности и содержанию действия. Выделяются следующие признаки юридического действия: их волевой характер, внешнее проявление воли лица, должно быть доступно восприятию со стороны третьих лиц, его формализация и его закрепление в норме права, направленность на правовые последствия. Выражение воли лица будет формироваться в момент начала использования системы ИИ, то есть лицо осознает и заранее нацелено на получение всех юридических последствий такого использования системы. При этом в этот же момент должна быть возможность у пользователя системы ИИ отказаться от принятия автоматизированного юридического решения в отношении него. Важной организационно-правовой, методологической и технологической задачей является такая организация функционирования алгоритмов, технологий и систем ИИ, чтобы любой значимый процесс принятия решения был прозрачен[4].

5. Важно отметить, что в частном праве в качестве обязательного условия наступления юридически значимых последствий является добросовестность участников отношений при создании и эксплуатации систем ИИ. Процесс принятия решения сосредоточен вокруг принимающего решение (*агента, человека или автоматизированной системы*), который делает выбор из доступных ему вариантов. Принимающий решение будет считаться разумным, если в основе его выбора лежит *механизм рассуждения*[5], то есть этот элемент является ключевым при квалификации определенного действия в качестве решения. Принятие решения относится к действиям лиц, то есть фактам, которые прямо зависят от воли и сознания людей. При этом если мы говорим о юридически значимых действиях, то воля должна быть направлена *на наступление правовых последствий*. Это в полной мере будет относиться и к ситуациям, когда принятие решения будет делегировано системе ИИ, можно предположить, что при создании системы ИИ воля и сознание людей будут *выражены в коде*.

6. В большинстве случаев ИИ определяется через сравнение со способностями человека или через возможности технологий обрабатывать данные и информацию. Понятие ИИ в России основывается на сравнении с когнитивными и интеллектуальными функциями человека. Под ИИ понимают комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека[6]. В США ИИ понимается как: система, выполняющая задачи в неопределенных условиях без

[4] См.более подробно раздел 5 настоящего доклада.
[5] A. Rosenfeld, S. Kraus. Predicting Human Decision-Making: From Prediction to Action. Morgan & Claypool, 2018, p.7.
[6] Национальная стратегия развития ИИ на период до 2030 года Указ Президента РФ 10 октября 2019 г. № 490.

контроля со стороны человека, или система, способная учиться на собственном опыте и совершенствовать выполнение; система, думающая как человек; система, действующая как человек; технологии, пытающиеся приблизится к выполнению некоторых когнитивных функций человека; системы, действующие рационально[7]. Департамент обороны США определяет ИИ как набор различных техник и технологий по обработке информации, используемых для выполнения целенаправленной задачи или как средство анализа для выполнения задачи[8].

7. Смежным с ИИ является понятие *системы с ИИ*. В этом случае ИИ позиционируется как компонент программного обеспечения, устройства или машины, который образует единую систему с ИИ. В США система с ИИ определяется как система, которая имеет в своем составе ИИ[9], или как техническая система, использующая ИИ для решения проблем[10]. Системы с ИИ по степени делегирования могут подразделяться на несколько категорий. Так, американское «Общество автомобильных инженеров» (SAE) выделяет шесть уровней автоматизации вождения автомобиля в диапазоне от нуля (полностью ручное управление) до пяти (полностью автономное), которые также обозначают степень делегирования[11]. Кроме этого, выделяют классификации ИИ по степени риска, способностям, категориям задач, эволюционному развитию и моральному отношению.

8. Наиболее распространенной классификацией ИИ является подразделение по способностям на «*Общий*» и «*Узкий*» (специальный).[12] В некоторых нормативных актах и литературе используются термины, соответственно, «*Сильный*» и «*Слабый*» ИИ. По степени риска возникновения значимых правовых последствий системы с ИИ делятся на системы с высокой и низкой степенью риска[13].

9. Правовые проблемы соответствия процессов разработки и применения ИИ правовым и этическим принципам включают в себя вопросы обеспечения справедливости, соблюдения прав человека, определение норм ответственности за причинение вреда, предупреждение дискриминации при использовании ИИ, соблюдение конфиденциальности персональных данных (далее – ПД), баланса между эффективностью и публичными интересами; иные

---

[7] FUTURE of Artificial Intelligence Act of 2017 H.R.4625 US https://www.congress.gov/115/bills/hr4625/BILLS-115hr4625ih.pdf.
[8] Board, D. I. (2019). AI principles: Recommendations on the ethical use of Artificial Intelligence by the Department of Defense. Supporting document, Defense Innovation Board.
[9] Board, D. I. (2019). AI principles: Recommendations on the ethical use of Artificial Intelligence by the Department of Defense. Supporting document, Defense Innovation Board.
[10] NIST (2019) U.S. LEADERSHIP IN AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools. https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf.
[11] SAE International, J3016_201806: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (Warrendale: SAE International, 15 June 2018).
[12] FUTURE of Artificial Intelligence Act of 2017 H.R.4625 US.
[13] European Commission. White paper on artificial intelligence–a European approach to excellence and trust. – 2020.

проблемы. Проблема баланса между эффективной разработкой технологии и публичными и частными интересами является одной из наиболее важных в рассматриваемой сфере. Установление процедур упрощенного тестирования и внедрения технологических решений, разработанных на основе ИИ, а также делегирование информационным системам, функционирующим на основе ИИ, возможности принятия отдельных решений могут противоречить правам человека на неприкосновенность частной жизни, личную и семейную тайну, защиту своей чести и доброго имени, на тайну переписки, телефонных переговоров, почтовых и иных сообщений.

10. ИИ в большой степени зависит от используемых при обучении и работе данных. Анализ больших объемов информации является необходимым условием работоспособности систем с ИИ. Искажение используемых ИИ данных происходит при некорректной разметке, использовании недостаточных или нерепрезентативных данных, искажении данных при получении с сенсоров. Кроме этого, одной из причин искажений может являться некорректная выборка данных, отражающая личные предпочтения («необъективность») разработчика. Решения, предлагаемые ИИ, в большинстве случаев *не являются полностью объяснимыми, предсказуемыми и прозрачными*. Функционирование ИИ происходит в режиме «черного ящика», при котором трудно установить причину конкретного решения. При использовании системы с ИИ проблема объяснимости и обоснованности предлагаемого ИИ решения приводит к невозможности объективной оценки пользователем последствий своих действий.

11. Проблема *агента* возникает в правовых отношениях, в которых ИИ действует *автономно*. Такие отношения отличаются взаимодействием субъектов через технического агента. При этом у субъектов правоотношений отсутствует возможность вмешаться в процесс исполнения задания, возможно только предварительное или последующее одобрение действий и решений, реализованных ИИ. Решение правовых проблем должно быть направлено на предупреждение негативных последствий при использовании систем с ИИ, устранение правовых неопределенностей с учетом снижения барьеров для участников рынка по внедрению технологии ИИ. Проблемы соответствия процессов разработки и применения ИИ для целей принятия юридически значимых решений правовым и этическим принципам включают в себя вопросы обеспечения справедливости, соблюдения прав человека, уточнение норм ответственности за причинение вреда, предупреждение дискриминации и соблюдение конфиденциальности ПД. Необходимо также обеспечить человеко-ориентированный подход к разработке, внедрению и использованию ИИ, включающий в себя обеспечение защиты

гарантированных российским и международным законодательством прав и свобод человека и повышение благосостояния и качества жизни граждан[14].

12. Одной из важнейших проблем является достижение баланса между экономическими интересами и интересами общества, ставшая отражением глубокого *конфликта выгоды и морали*. С позиции теории, наибольшую значимость имеют следующие ключевые вопросы в области использования ИИ: проблема этики самого ИИ, возможности применения этой категории к ИИ и возможности программно заложить этические нормы в ИИ; проблема этического взаимодействия системы «человек – ИИ», включая проблематику создания этических основ при организации машинного обучения систем ИИ. Группа экспертов по искусственному интеллекту Европейской комиссии в 2019 году приняла «Этические принципы надежного искусственного интеллекта»[15]. Надежный ИИ имеет три компонента, которые должны соблюдаться на протяжении всего жизненного цикла системы:

1) он должен соответствовать требованиям закона;

2) он должен обеспечивать соблюдение этических принципов и ценностей;

3) он должен быть надежным как с технической, так и с общественной точки зрения, поскольку даже при благих намерениях системы ИИ могут причинить непреднамеренный вред.

13. В некоторых этических принципах существует разное отношение к негативным последствиям. Одни стандарты рекомендуют избегать рисков, другие концентрируются на ответственности за риски. Так в стандартах, разработанных Институтом инженеров по электротехнике и электронике (IEEE)[16], указывается на необходимость избегать неправильного использования ИИ, а в Монреальской декларации[17] утверждается, что разработчики ИИ должны взять на себя свою ответственность за риски, связанные с осуществлением ими технологических инноваций. Рекомендуемыми базовыми этическими основаниями являются: соблюдение прав и свобод человека, недопущение дискриминации, политическая нейтральность, справедливость и честность, благополучие окружающей среды, прозрачность и объяснимость ИИ, раскрытие информации.

14. Одной из центральных проблем является *отсутствие достаточной степени доверия* со стороны общества к использованию цифровых технологий, в частности искусственного интеллекта и робототехники. Корни проблемы выстраивания доверия к новым

---

[14] Концепция развития регулирования отношений в сфере технологий ИИ и робототехники до 2024 года. Утверждена распоряжением Правительства Российской Федерации от 19 августа 2020 г. N 2129-р.

[15] HLEG A. I. High-level expert group on artificial intelligence //Ethics Guidelines for Trustworthy AI. – 2019.

[16] IEEE, Global Initiative et al. Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems. – 2018.

[17] The Montréal Declaration for a Responsible Development of Artificial Intelligence. URL: https://www.montrealdeclaration-responsibleai.com/the-declaration.

технологиям уходят к понятию «черный ящик», которым сегодня часто описывается состояние, когда человек не понимает, как именно работает та или иная технология. В России описываемая проблема является особенно сложной, поскольку недоверие к технологиям часто связано с недоверием к органам власти. Одним из важнейших направлений в решении описанной проблемы является систематическое изучение социокультурной стороны процессов цифровой трансформации и, в частности, этических вопросов разработки и развития цифровых технологий, следствием чего станет выработка соответствующих рекомендаций, касающихся нормативно-правового регулирования технологий и работы с общественным мнением.

15. В России этические проблемы развития и использования цифровых технологий в РФ не являются на данный момент действительно важной частью повестки не только государственных органов власти и коммерческих структур, но и общества в целом, поскольку:

- в России пока не сформирован полноценный институт репутации;

- в российском обществе запрос на защиту прав и свобод на данный момент не является центральным;

- научное и экспертное сообщество не обладает достаточным влиянием ни на процессы принятия решений государственных институтов и коммерческих структур, ни на общественное мнение.

16. В *Европейском союзе* тема принятия ИИ юридически значимых решений широко развита. Действует большое количество документов, посвященных рискам, рекомендациям и принципам в сфере ИИ, в том числе применительно к принятию автоматизированных решений. Среди них – Сообщение Комиссии Европейского парламента «ИИ для Европы», Декларация о кооперации по вопросам ИИ в ЕС, Объединенный план по содействию развития и использования ИИ в Европе, Резолюция Европейского парламента с рекомендациями Комиссии по вопросам гражданско-правового регулирования робототехники, Открытая концепция по ИИ и другие. В европейском дискурсе принятие делегирования решений системам ИИ рассматривается, прежде всего, в контексте потенциальных угроз для прав и свобод человека, риска дискриминации и др. В связи с чем, для минимизации рисков, которые несет делегирование принятия решений ИИ, разработано значительное число документов с рекомендациями и принципами, включая вопросы этики. ЕС осознает важность создания благоприятной правовой и социальной среды для развития технологий в конкретной юрисдикции.

17. В *США* тема принятия автоматизированных решений также достаточно проработана. Так, в контексте кредитного скоринга действует Закон о беспристрастной кредитной отчетности (*Fair Credit Reporting Act*), а в рамках скоринга благонадежности

потенциальных арендаторов – Закон о справедливом решении жилищных вопросов (*Fair Housing Act*), Закон штата Иллинойс о прозрачности, согласии и уничтожении данных в связи с проведением видео-интервью с использованием ИИ. Выявлены общие черты регулирования: запрет на дискриминацию при принятии решений, требования к прозрачности процесса принятия решений, обязанность предварительного информирования пользователя о взаимодействии с ИИ. Существуют инициативы по внедрению комплексного регулирования, например, Закон об автоматизированной системе принятия решений от 2020 года Калифорнии и Законопроект 2019 года о прозрачности алгоритмов (*Algorithmic Accountability Act*). При этом принятие юридически значимых решений системами ИИ в США детально не урегулировано. Наряду с развитием и поощрением использования систем ИИ при принятии определенных решений, в США существуют тенденции и по ограничению использования систем ИИ в отдельных сферах. Наиболее ярко данная тенденция выражается в ведении моратория на использование технологий распознавания лиц.

18. Собственно, подобное отсутствие подробного регулирования характерно и для ситуации в мире в целом. Сложность формирования регулирования в данной сфере заключается в том, что есть «*риск того, что выявлены не все риски*». При этом предварительное, проактивное регулирование показывает свою неэффективность, если такие риски изучены не до конца и нет полного понимания направлений развития технологий. В этой связи предлагается разрабатывать регулирование, снижающее *очевидно существующие* риски, и при этом вести постоянную аналитическую работу над дополнительным изменением регулирования. Как правило, уполномоченными органами в той или иной сфере разрабатываются рекомендации, программные документы и документы стратегического характера.

19. Среди сфер, в которых возможно и фактически осуществляется или планируется осуществлять делегирование полномочий, функций, обязанностей системам ИИ, с учетом проведенного анализа возможно выделить следующие: государственные услуги и управление; правоохранительная деятельность (распознавание лиц, назначение штрафов, обработка данных в уголовных делах); судебная деятельность (поддержка решений судьи, обеспечение единообразия судебной практики, скоринговые системы, оценивающие вероятность совершения рецидива, рассмотрение определенных категорий споров); таможенная деятельность; сфера здравоохранения; автоматизированный транспорт (беспилотный транспорт, использование дронов для проверки трубопроводов); экология; банковская деятельность (кредитный скоринг, предложение персонализированных услуг); подбор персонала; рынок недвижимости (скоринг благонадежности арендаторов).

20. Среди ключевых требований к системам ИИ, которым могут быть делегированы

какие-либо полномочия, функции, обязанности, можно привести следующие: отслеживаемость, объяснимость, прозрачность и проверяемость; подотчетность и ответственность за действия таких алгоритмов, эффективность решений, принимаемых алгоритмами; отсутствие необъективности, дискриминации, неравенства, несправедливых решений, порождаемых алгоритмами, предвзятости алгоритмов; соответствие систем ИИ не только нормам права, но и этики; возможность контроля со стороны человека; устойчивость и безопасность. Отмечается, что чем меньше человеческого контроля предполагает конкретный случай делегирования полномочий системам ИИ, тем более жесткие требования должны предъявляться к тестированию и системному управлению технологиями ИИ.

21. Основным барьером для организации обучения систем ИИ является регулирование в сфере ПД. В связи с этим на этапах доступа, передачи и иной обработки ПД для обучения систем ИИ должны соблюдать права субъектов ПД. Еще одним барьером является ограниченность набора данных для обучения, которая увеличивает риски дискриминации и предвзятости при принятии решений. Для снижения рисков дискриминации их необходимо учитывать в самом начале обучения – на этапе формирования тренировочного набора, кроме того, выходом может являться аудит (тестирование) системы, оценка влияния ИИ на защиту данных, обеспечение доступа к коду программного обеспечения. Высказываются дискуссионные позиции, согласно которым необходимо внесение изменений в законодательство, которые должны позволить легально использовать большие объемы данных, поскольку предвзятость алгоритмов связана в том числе с ограниченностью данных, используемых в тренировочных наборах.

22. На текущий момент речь в основном не идет о возможности делегирования системам ИИ принятия окончательных решений, а также о полном исключении субъекта из процесса принятия решения. Такая предосторожность связана с тем, что существуют риски необъективности ИИ. С одной стороны, может казаться, что ИИ может позволить избежать недостатков, присущих человеческому мышлению, однако с другой стороны ИИ может быть обучен на необъективных данных и суждениях. На уровне Европейского союза речь идет о следовании «человекоцентричному» подходу, согласно которому человек должен гарантированно получать контроль над ИИ в любой момент времени. Важным элементом проектируемой концепции регулирования ИИ в ЕС является риск-ориентированный подход. Применение ИИ будет отнесено к категории высокорисковых в случае отнесения ИИ к определенной сфере или способу использования, а также исходя из уровня влияния решений на права граждан. Также предлагается установить приемлемый уровень пересмотра человеком решений и действий ИИ. При этом есть примеры, когда ИИ может быть полностью

делегировано принятие не юридически важных решений, но иных решений (в частности, беспилотный транспорт).

23. Что касается последствий принятия решений ИИ, то отмечается, что физическое лицо должно иметь право на оспаривание решения, принятого системами ИИ. Специальный порядок оспаривания в нормативных правовых актах исследуемых юрисдикций не предусмотрен. Однако само право на обжалование, например, по умолчанию предоставляется ч. 3 ст. 22 GDPR. Возможность обжалования решений предусмотрена Директивой по автоматизированному принятию решения Канады, в рамках системы социального рейтинга Китая, а также рекомендуется в этических рекомендациях в рамках программы Smart Dubai и предполагается предусмотреть в Эстонии в рамках проекта, связанного с внедрением робота-судьи. В то же время встречаются и случаи, когда обжалование решений фактически является невозможным (Закон о беспристрастной кредитной отчетности США).

24. Делегирование принятия решения ИИ может требовать получения согласия физического лица, в отношении которого принимается такое решение. Необходимость получения согласия может проистекать из регулирования ПД (например, в GDPR). Среди дополнительных прав субъектов, в отношении которых системами ИИ принимаются решения, стоит отметить следующие: право знать, что решение принимается ИИ; право иметь возможность контролировать генерируемые в процессе использования ИИ данные и знать, куда в дальнейшем они направляются и как взаимодействуют с устройствами и с другими людьми; право быть уведомленным о том, как связаться с человеком и как удостовериться в том, что принимаемые системой решения могут быть проверены или скорректированы; право на получение объяснений. Данное требование может быть трудновыполнимым в случае с ИИ, чью логику решения не всегда могут описать даже разработчики.

25. Разработка специальных принципов правового регулирования делегирования принятия юридически значимых решений с использованием систем ИИ должна учитывать и основываться на *общих принципах справедливости, разумности и добросовестности*. Вместе с тем, в связи с особенностями использования технологий ИИ, например, справедливость наполняется новым содержанием, которое охватывает, в том числе, соблюдение общечеловеческих ценностей, равный и недискриминационный подход, социальную справедливость, прав человека, включая свободу, достоинство, независимость, защиту частной жизни, защиту ПД. Данные принципы могут рассматриваться как «нормативное ядро» принципиального подхода к этике и управлению в сфере ИИ с поправкой на разрыв между концептуальными формулировками и реальной имплементацией, а также на возможные различия в отдельных ценностных подходах в зависимости от юрисдикции.

26. При делегировании принятия решений ключевыми проблемами ответственности, выходящими на первый план, становятся *определение субъекта ответственности*, определение случаев наступления такой ответственности и определение вида ответственности. Праву известны ситуации, когда ответственность за действия одного субъекта перелагается на другого субъекта. К таким ситуациям, например, относятся ответственность работодателя за действия работника при исполнении им трудовых функций, ответственность родителей за действия несовершеннолетних детей, ответственность юридического лица или государственного органа за действия его должностных лиц, ответственность владельца источника повышенной опасности.

27. При рассмотрении ответственности ИИ целесообразно говорить, в первую очередь, о *деликтной* ответственности, то есть меры ответственности должны быть установлены как реакция на вред, который ИИ может причинить или причиняет. При этом речь не всегда идет о линейной ответственности, то есть ответственности одного лица за вред, который он причинил, а скорее о совмещенной ответственности, то есть, когда помимо причинителя вреда к ответственности могут быть призваны и другие субъекты[18]. В резолюции Европарламента от 16.02.2017[19] о применении норм гражданского законодательства к роботам определены две модели ответственности, которые потенциально могут быть использованы для определения ответственности робота: ответственность независимо от вины, рискориентированный подход, когда ответственным признается то лицо, которое способно было минимизировать риск. 20 октября 2020 года Европарламентом были приняты несколько резолюций, одна из которых касается режимов гражданской ответственности для ИИ[20]. В ней, в частности, отмечается, что все физические или виртуальные действия, устройства или процессы, которые управляются системами ИИ, технически могут быть прямой или косвенной причиной вреда или ущерба, но почти всегда оказываются результатом создания, развертывания или вмешательства в системы.

28. В случае признания ИИ объектом также есть несколько вариантов возложения деликтной ответственности:

1) на *обладателя прав* на устройство, снабженное интеллектом;

2) на *разработчика* программного обеспечения;

3) на *оператора*, обслуживающего ИИ.

---

[18] См. Тихомиров Ю.А, Крысенкова Н.Б., Нанба С.Б., Маргушева Ж.А. Робот и человек: новое партнерство? // Журнал зарубежного законодательства и сравнительного правоведения. 2018. №5. С. 5- 10.

[19] Civil Law Rules on Robotics European Parliament resolution of 16.02.2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). EP reference number: P8_TA (2017)005.

[20] Civil liability regime for artificial intelligence. European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL))URL: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.pdf

29. Наличие собственных воли и интереса не всегда отличают субъекта от объекта, но чаще всего являются необходимыми элементами для применения норм ответственности. ИИ в случае, когда он наделен собственной волей и интересом, должен рассматриваться как самостоятельный субъект ответственности, в том случае, если он не наделен собственными волей и интересом, его можно рассматривать как агента. Ключевым является определение статуса ИИ. Только определив и закрепив статус ИИ, можно говорить о вопросах ответственности. Представляется, что она должна строиться с учетом роли и доктрины всех институтов регулирования. В случае признания ИИ только объектом права, к чему пока что склоняются в мире, ответственность должна быть распределена по разным этапам жизненного цикла ИИ (этап разработки, этап эксплуатации, этап утилизации). При этом ответственность ИИ сводится не к карательно-воспитательным мерам, а к установлению действенного механизма управления рисками.

30. В свете возможного возложения ответственности на разработчиков необходимо хотя бы на начальных этапах предусмотреть сбалансированную систему иммунитетов для них, добавив обязательное *страхование ответственности*, а также *регистрацию систем ИИ*. В случае признания ИИ субъектом возможно установление режима совмещенной ответственности, когда субсидиарную ответственность могут нести и создатель ИИ, и его владелец или иной субъект.

31. *Ценности* являются основным определяющим фактором при построении концепции правового регулирования ИИ. Разработка и применение ИИ должны стремиться к увеличению благосостояния общества и отдельного человека с сохранением гарантированных прав и свобод человека и обеспечением безопасности технологии ИИ для человека. Само по себе благосостояние не имеет ценности без соблюдения этических норм, влияющих на формирование права. При формировании конкретных правовых норм концепция должна не допускать дискриминации и позволять соблюдать баланс между интересами отдельных групп людей, отраслей промышленности и национальными интересами. Отношения не только между людьми, но и человека с ИИ *должны быть справедливыми*. В частности, при регулировании отношений в сфере разработки систем с ИИ необходимо обеспечить соблюдение каждой системой ИИ интересов и учет особенностей правового статуса детей, инвалидов и пожилых людей. Дискриминация, доминирование и манипуляция со стороны ИИ недопустимы. Решения ИИ, затрагивающие права человека, должны быть прозрачными для того, чтобы человек мог проверить их на объективность, логику, достоверность и иметь право обжаловать данные решения в случае дискриминации. Информация о том, какие персональные данные используются ИИ при взаимодействии с человеком, должна быть предоставлена этому человеку по его требованию. Право на получение такой информации

должно быть разъяснено доступным и понятным способом до начала взаимодействия человека с ИИ. Решения ИИ не могут приниматься как истинные пока не объяснены. Существует система с ИИ, чьи решения не являются объяснимыми. Нетранспарентные автономные системы ИИ, генерирующие непредсказуемые решения, затрагивающие права человека, не могут применяться без специальных методов контроля, подтверждения или создания специальных условий их использования.

32. Помимо этических принципов и законодательных реформ, необходимо уделить внимание *стандартам*. В таких сферах как промышленность и транспорт это должны быть стандарты требований безопасности. Во всех других – стандарты хранения и передачи данных. Помимо разработки новых стандартов необходимо пересмотреть текущие стандарты в разных отраслях на применимость к ИИ. Также в стандартах может быть закреплен следующие метод: обязательность технического внедрения концепции «отказа от предсказания» ИИ. Данная концепция означает, что, если система ИИ «не уверена» в принимаемом решении, она должна иметь возможность не принимать его/не совершать какое-либо действие, либо как минимум уведомлять человека об уровне «уверенности» в том или ином решении или действии.

33. В число предложений по регулированию вопросов принятия юридически значимых решений системами ИИ в отдельных сферах входит способ закрепления сфер, в которых *допустимо использование систем И*И при принятии юридически значимых решений. Первый способ – закрепление конкретных сфер, в которых будет допустимо использовать ИИ для принятия юридически значимых решений. На сегодня можно сказать, что принятие решений ИИ допустимо, как минимум, в следующих сферах: государственные услуги и управление, правоохранительная деятельность, судопроизводство, таможенная деятельность, сфера здравоохранения, транспорт, банковско-кредитная сфера, подбор персонала и недвижимости, ритейл и «цифровые услуги» (B2B и B2C), игровая индустрия и киберспорт. При этом может быть установлено требование по регулярному обновлению перечня. Второй способ – закрепление лишь некоторых принципов, на основе которых будет строиться регулирование использования систем ИИ при принятии юридически значимых решений. Такой подход связан с тем, что фиксация конкретного списка допустимых и недопустимых случаев использования систем ИИ для каждой сферы может быть затруднительным. Помимо принципов также важно определить степень и меры ответственности за вред, причинный системами ИИ при принятии юридически значимых решений и в иных случаях. Сферы и конкретные случаи допустимости/недопустимости использования ИИ при принятии юридически значимых решений могут быть определены постфактум. Представляется целесообразным реализация *смешанного подхода*, который предполагает как закрепление конкретных сфер, так и закрепление принципов.

Регулирование ответственности в сфере искусственного интеллекта

34. Формирование института юридической ответственности в сфере создания и использования систем и технологий искусственного интеллекта (ИИ) является одной из важных задач, стоящих перед правом в XXI веке. Растущая автономность действий систем ИИ и снижение степени участия человека ставит вопрос о достаточности существующих правил регулирования в отношении института юридической ответственности, включая гражданско-правовую, административную, уголовную ответственность.

Концепция развития регулирования в сфере технологий искусственного интеллекта (ИИ) и робототехники (РТ), утвержденная распоряжением Правительства РФ от 19.08.2020 № 2129-р[21], определяет, что реальный уровень развития технологий ИИ и робототехники (РТ) не предполагает кардинальных изменений в регулировании института юридической ответственности, однако требует постепенной доработки его отдельных элементов. Требуется дальнейшая проработка механизмов гражданско-правовой, уголовной, административной ответственности в случае причинения вреда системами ИИ и РТ, имеющими высокую степень автономности при принятии ими решений, в том числе с точки зрения определения лиц, которые будут нести ответственность за их действия, доработки, при необходимости, механизмов безвиновной гражданско-правовой ответственности, а также возможности использования способов, позволяющих возместить причиненный действиями систем ИИ и роботами вред (например, страхование ответственности, создание компенсационных фондов и т.д.). Также при наличии реального риска нарушения прав и свобод граждан может быть актуальной проработка вопроса об условиях самоидентификации системы ИИ при прямом взаимодействии с человеком. Общий вектор возможных изменений должен быть направлен на то, чтобы гарантировать эффективное и справедливое функционирование институтов юридической ответственности и распределение ответственности в случае такого причинения вреда. При этом справедливое отношение должно вырабатываться с учетом того, что потенциальные объемы ущерба, который один может причинить другим, значительно выросли из-за уровня связности субъектов оборота, который является обычным явлением в формирующейся цифровой экономике.

35. Согласно Национальной стратегии развития искусственного интеллекта на период до 2030 года, утвержденной указом Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (далее Национальная стратегия развития ИИ), под искусственным интеллектом понимается комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение

[21] Доступна: https://www.tadviser.ru/images/c/cf/Концепция_ИИ_и_РТ.pdf

и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе, в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений (п.5). Данное определение позволяет сделать вывод о том, что системы ИИ могут не иметь какого-либо вещественного выражения (детальное исследование необходимо проводить с учетом деления систем ИИ по критерию тоже). Круг субъектов, которые будут задействованы на каждом этапе жизнедеятельности системы ИИ, видимо, будет различаться. Следуя вышеназванным двум категориям ИИ, в исследованиях, как правило, речь идет о классификации ИИ и делении его на так называемый «сильный ИИ» и «слабый ИИ». Предполагается, что данное деление имеет значение для выбора модели регулирования в зависимости от того, насколько автономно действует система ИИ.

36. Применительно к системам ИИ в правовых исследованиях главный акцент делается на вопросах гражданско-правовой ответственности. Хотя, безусловно, значимость проведения в том числе и уголовно-правовых исследований очевидна[22], поскольку, если верить прогнозам, возможность создания искусственного интеллекта, который можно было бы сравнить с человеческим интеллектом или превзойти его, вполне реальна и может быть достигнута в следующих десятилетиях. В этой связи ученые оценивают перспективу применения к ИИ мер уголовно-правового характера и появление в уголовных кодексах раздела, посвященного мерам уголовно-правового характера для электронных субъектов[23].

37. Гражданское законодательство допускает применение мер ответственности в отсутствие вины, если такая возможность предусмотрена законом или договором и также позволяет применять различные формы ответственности, такие как смешанная ответственность при совместном причинении вреда; ответственность за вред, причиненный деятельностью, создающей повышенную опасность для окружающих; ответственность производителя за качество товара (услуги) − ответственность перед потребителем. Указанные юридические конструкции могут быть использованы при установлении правового регулирования отношений с участием систем ИИ. Традиционно выделяют преддоговорную, договорную и внедоговорную гражданско-правовую ответственность.

---

[22] Radutniy O. E. Criminal liability of the artificial intelligence. doi: 10.21564/2414-990x.138.105661 UDC 343.22+343.412:004.056//cyberleninka.ru/article/n/criminal-liability-of-the-artificial-intelligence/viewer; Грачева Ю.В., Арямов А.А. Роботизация и искусственный интеллект: уголовно-правовые риски в сфере общественной безопасности // Актуальные проблемы российского права. 2020, № 6// СПС КонсультантПлюс; Некрасов В.Н. Актуальные вопросы уголовно-правовой охраны инновационной деятельности в России // Актуальные проблемы российского права. 2017, № 7// СПС КонсультантПлюс.

[23] Radutniy O. E. Criminal liability of the artificial intelligence.doi: 10.21564/2414-990x.138.105661 UDC 343.22+343.412:004.056//cyberleninka.ru/article/n/criminal-liability-of-the-artificial-intelligence/viewer

38. Для внедоговорной ответственности, которая развивается в настоящее время применительно к ИИ в целях лучшего понимания предсказуемости и установления причинно-следственной связи в сложных сценариях[24], можно выделить следующие особые характеристики случаев причинения вреда системами ИИ: делегирование системе ИИ самостоятельного участия в обороте, в том числе принятия решений (при неопределенности правового положения системы ИИ); непрозрачность принятия решения, и как следствие, его непредсказуемость. В этой связи возможно сформулировать самым общим образом три вида фактических оснований, которые могут быть достаточными для привлечения к ответственности[25]: принятие недостаточных мер предосторожности при создании (проектировании и производстве) системы, тем самым оставляя риски, которые можно было бы предотвратить, принятие недостаточных мер предосторожности или уделение недостаточного внимания при владении и использовании системы (например, может включать в себя отказ от обновления системы), риск, то есть возникновение ответственности, основанной на риске.

39. Исследование мирового опыта регулирования общественных отношений, связанных с вопросами ответственности в сфере ИИ, позволяет сделать вывод о том, что по наличию правового регулирования ответственности в сфере ИИ все страны делятся на три большие группы:

1) страны, в которых специальное регулирование ответственности за вред, причиненный системами ИИ, отсутствует в целом. Это подавляющее большинство исследованных стран;

2) страны, в которых регулирование ответственности за вред, причиненный системами ИИ, специальным образом сформулировано применительно к высокоавтоматизированному (беспилотному) транспорту;

3) страны, в которых регулирование в сфере беспилотного транспорта не содержит информацию об ответственности за вред, причиненный системами ИИ, но устанавливает требование о страховании ответственности.

40. В тех случаях, когда сложно определить ответственное лицо, либо, понимая некоторую несправедливость возложения ответственности на физическое лицо, когда вред был причинен автомобилем в режиме автопилотирования, страны вводят модель

[24] Lior, Anat. The AI Accident Network: Artificial Intelligence Liability Meets Network Theory (March 26, 2020). Work published in 95 TUL. L. REV. (2020-2021)// https://ssrn.com/abstract=3561948

[25] Tjong Tjin Tai, Eric, Liability for (Semi)Autonomous Systems: Robots and Algorithms (April 13, 2018). Vanessa Mak, Eric Tjong Tjin Tai and Anna Berlee (eds), Research Handbook on Data Science and Law (Edward Elgar, 2018), p. 55-82, Tilburg Private Law Working Paper Series No. 08/2018 , Tilburg Law School Research Paper No. 2018-9. Доступно: https://ssrn.com/abstract=3161962 или http://dx.doi.org/10.2139/ssrn.3161962

обязательного страхования. Такая модель позволяет компенсировать ущерб, причиненный пострадавшей стороне. При этом отмечается необходимость разработки соответствующего регулирования. Наиболее проработанная концепция возможного подхода к ответственности за вред, причиненный ИИ, на сегодня разработана в ЕС. В частности, предлагается дифференцированный подход к ответственности, зависящий от типа ИИ: *безвиновная* ответственность для *высокорисковых* систем ИИ и *виновная* – для *всех иных*.

41. По критерию возможности применения имеющегося нормативного правового регулирования к случаям вреда, причиненного системами ИИ возможны следующие подходы:

- ответственность за вред, причиненный *источником повышенной опасности*;

- ответственность за причинение вреда *вследствие недостатков (дефектов) продукта*.

- безвиновная ответственность за вред, причиненный *чрезвычайно опасной деятельностью*.

- применение по аналогии норм об ответственности *за вред, причиненный животными*. В частности, между роботами и животными можно обнаружить некоторое сходство. Например, и роботы. и животные могут действовать независимо от своих владельцев, воспринимать окружающую обстановку и осуществлять действия в зависимости от нее.

- применение по аналогии норм об ответственности *за вред, причиненный работниками*. Ответственность работодателя за вред, причиненный работником третьим лицам, связана с действиями работника, повлекшими причинение вреда, которые он совершил в пределах выполнения своих рабочих обязанностей.

- применение по аналогии норм об ответственности *за вред, причиненный детьми*. Сравнение систем ИИ с детьми основано на том, что их действия могут быть относительно непредсказуемыми. Кроме того, отмечается, что как дети подвержены влиянию и воспитанию со стороны родителей, иных третьих лиц, так и системы ИИ могут подвергаться влиянию или обучаться. В то же время, уровень влияния, которое разработчики могут оказываться на системы ИИ, может быть ограниченным.

- применение *договорной ответственности*.

42. Что касается уголовной ответственности, то ответственность может лежать только на человеке, поскольку наделение систем ИИ правосубъектностью нецелесообразно на данном этапе. Это связано с тем, что действующее уголовно-правовое регулирование в целом «человекоцентрично» и не предполагает возложения ответственности не на человека.

43. Примеры реальных случаев причинения вреда системами ИИ на сегодня немногочисленны, однако встречаются. В качестве наиболее показательного примера стоит

привести дело, связанное с аварией высокоавтоматизированного автомобиля компании Uber, который, двигаясь в автономном режиме, допустил наезд на пешехода в темное время суток, в результате которого пешеход скончался. Данное дело представляет интерес правовыми последствиями для компании Uber и водителя (оператора). На компанию Uber не были возложены какие-либо меры ответственности, однако в отношении оператора в настоящее время ведется расследование. Указанный пример может являться одним из примеров возможных подходов к разрешению вопросов ответственности за вред, причиненный системами ИИ. Так, например, если система ИИ является контролируемой и человек обязан осуществлять контроль ее действий, то в случае причинения вреда в результате того, что оператор такой контроль не осуществлял, а при должном контроле мог бы предотвратить вред, допустимо возлагать ответственность именно на оператора, а не производителя, разработчика или иных лиц, участвовавших в создании системы ИИ.

44. Обобщение и систематизация принципов правового регулирования ответственности в сфере ИИ, следующих как из сравнительно правового, так и из иных составляющих настоящего исследования, позволяет построить следующую систему принципов для последующей их реализации на различных уровнях системы российского права:

- ответственность за причинение вреда системами с ИИ должны нести «*традиционные*» субъекты права.

- основным ответственным должно являться то лицо, с кем ассоциирован *риск операционного использования, или лицо, являющееся непосредственным выгодоприобретателем* от использования системы с ИИ.

- пользователь системы с ИИ должен нести *ответственность за выбор технологии*, не соответствующей задаче, за нарушение правил по использованию, контролю и техническому обслуживанию системы.

- производитель систем с ИИ должен нести *ответственность за вред, причиненный недостатками системы* с ИИ, полученными во время производства системы.

- разработчик и производитель систем с ИИ должен обеспечивать *соответствие проекта и правил разработки задачам*, полное раскрытие информации о системе с ИИ (но при этом – с учетом баланса интересов в части интеллектуальных прав правообладателя с точки зрения авторского права и возможного режима конфиденциальности отдельных данных с точки зрения прав обладателя информации, а также в целях обеспечения безопасности систем с ИИ), недопустимость рекламы или распространения иной информации и системе с ИИ, вводящей в заблуждение потенциальных пользователей.

- за исключением случаев, когда это прямо следует из научно-обоснованной междисциплинарной (включая, но не ограничиваясь этим, техническую, социальную, экономическую и юридическую составляющие) оценки рисков, предупредительные *регуляторные меры, направленные на снижение риска, не должны ограничивать развитие технологии ИИ и не должны идти вразрез с регуляторными мерами стимулирующего характера*.

- при использовании систем с ИИ возможно применение *ответственности без вины* (источник повышенной опасности) в тех случаях, когда такая система с ИИ может квалифицироваться как источник повышенной опасности и когда это не противоречит природе отношений.

- высокая степень автономности ИИ *не может служить основанием для уменьшения ответственности* разработчиков, производителей.

- если разработчик ИИ обладает большей степенью контроля над функционированием системы с ИИ, чем производитель, собственник или пользователь данной системы, это должно *увеличивать ответственность* разработчика за причинение вреда. Данный принцип может быть представлен в более универсальной интерпретации: *степень контроля над функционированием системы ИИ пропорциональна ответственности за причинение вреда* (если производитель ИИ обладает большей степенью контроля над функционированием системы ИИ, соответственно, это должно увеличивать ответственность производителя и т.п.).

- технологии должны предусматривать *доступ к операционным данным системы* с ИИ, которые позволяют установить причину инцидента и ответственное лицо. Уничтожение таких данных должно рассматриваться как правонарушение вплоть до уголовного.

- невозможность доступа к операционным данным системы с ИИ обусловливает трактовку обстоятельства дела *в пользу потерпевшего*.

- лицо, изменившее или повредившие данные или алгоритмы, необходимые для стабильной работы системы с ИИ, несет ответственность *за ненадлежащее функционирование системы*.

- регулирование ответственности должно строиться *на риск-ориентированном подходе*, который предполагает создание методики оценки рисков в целях ограничения существенных рисков при применении технологии ИИ и возмещения вреда при страховании ответственности.

- чем более сложной становится структура взаимоотношений между вовлеченными в контроль над технологией лицами, тем сложнее определить одно виновное в правонарушении лицо. Подход, при котором каждый раз необходимо определять ответственное лицо, является

недостаточной мерой по защите прав потерпевшего. Институт *обязательного страхования гражданской ответственности* даст лучшую защиту и доступ к компенсации.

- лицо, контролирующее использование системы с ИИ *на постоянной основе, должно нести обязанность* по страхованию.

- продукты и услуги с невысоким уровнем риска *могут быть предметом саморегулирования*, с принятием участниками внутренних стандартов производства и кодексов лучших практик. С повышением уровня возможного риска уровень регулирования должен повышаться вплоть до ограничения использования технологии.

- санкции и объем ответственности «традиционных» субъектов права *должны определяться человеком*. Правоприменение, исключающее участие человека, *недопустимо*.

Для страхования гражданской ответственности необходимо установить все факторы и обстоятельства, влияющие на риск причинения вреда и возникновение ответственности, в том числе: степень влияния действий разработчика, поставщика баз данных, производителя, законного владельца, оператора и пользователя на совершение системой с ИИ определенных действий или генерацию решений, уровень автономности системы ИИ, степень контроля указанных ранее лиц за действиями решениями системы и связанные с этим их обязанности.

45. Правовыми основами страхования гражданской ответственности за вред, причиненный системами с ИИ, являются:

- *Баланс интересов*. Необходимо соблюдать баланс между защитой общественных интересов и технологическим развитием.

- *Защита интересов страховых компаний*. Технология ИИ является новым нестандартизированным продуктом с значительными вариациями риска, который необходимо определять в отсутствие исторической базы данных вероятного вреда, наносимого такими системами. Для выработки оптимального подхода к оценке рисков необходимо формирование единой базы страховых случаев с системами с ИИ и введение новых методов оценки рисков на основе онлайн мониторинга данных, экспертных мнений или зарубежных данных по страховым рискам.

- Страхование гражданской ответственности как *условие допуска на рынок*. В условиях стимулирования развития технологии и снятия правовых барьеров по допуску технологии к общественному использованию, требование об обязательном страховании является главным инструментом смягчения рисков использования технологии ИИ.

- *Определение ответственного лица*. Необходимо совершенствование текущих правил определения ответственного лица.

- Оценка риска *в условиях неполной информации*. Одной из проблем оценки страхового риска ответственности за вред, причиненный системами ИИ, является сложность

установления баланса между точностью оценки, с одной стороны, и защитой интеллектуальных прав. Необходимо разработать механизмы, позволяющие страховщикам оценивать страховые риски в отсутствие детальных сведений об устройстве системы ИИ.

- *Взаимосвязь страхования, сертификации и технического осмотра*. До момента страхования гражданской ответственности за вред, причиненный системами с ИИ, необходимо выполнять требования по постановке на государственный учет, сертификации и техническому осмотру объектов, оснащенных системами ИИ, если к ним применяются такие требования. Требования, устанавливаемые для объектов, оснащенных системами ИИ, не могут быть ниже, чем требования, установленные для объектов, управляемых человеком.

- Определение *максимального размера страхового возмещения*. В условиях неопределенности в оценке рисков, ограничение максимального размера выплаты по страховому договору будет являться инструментом, позволяющим страховым компаниям снизить собственные риски при страховании гражданской ответственности за вред, причиненный системами с ИИ.

- Страхование *изменчивой системы* с ИИ. Изменчивость ИИ делает возможным появление недостатков товара после продажи в течение срока службы товара. Необходимо определить случаи ограничения возмещения на выплаты по страховому договору, если изменения были произведены неправомерно.

- *Обязательное страхование*. Институт обязательного страхования гражданской ответственности должен дать потерпевшим оптимальную защиту и доступ к компенсации. Во избежание увеличения экономических затрат при внедрении ИИ, обязательное страхование не должно подменять страхование индивидуальных рисков и должно ограничиваться лишь сферами использования ИИ, имеющими высокий риск нанесения ущерба.

- *Компенсационный фонд*. При введении обязательного страхования гражданской ответственности рекомендуется сформировать компенсационный фонд для возмещения ущерба. Компенсационные фонды могут использоваться для помощи пострадавшим, которые имеют право на компенсацию в соответствии с применимыми правилами ответственности, но чьи требования по разным причинам не могут быть удовлетворены.

46. Особенности *гражданско-правовой ответственности* разработчика определяются, в том числе, тем обстоятельством, что функционирование системы ИИ, в результате которого был причинен вред, не обязательно подразумевает вину разработчика, но зависит от задач применения и обстановки применения, которые могут не совпадать с ограничениями, которые принимались во внимание при разработке системы. Ответственность производителя, в свою очередь, определяется возможными производственными дефектами, к которым разработчик может не иметь отношения. На уровне разработчика и производителя актуально техническое

регулирование, сертификация и, возможно, регистрация систем ИИ. Ответственность как производителя, так и продавца определяется главным образом двумя обстоятельствами: предоставлением достаточного объема информации о технических характеристиках системы ИИ и качеством товара, которое может быть ненадлежащим в том смысле, которое придается этому слову законодательством и практикой о защите прав потребителей. Ответственность владельца и пользователя системы (эти понятия следует разделять) особенно остро встает в отношении тех систем ИИ, которые отвечают признакам источников повышенной опасности. В целом также развитие гражданского законодательства и практики должно быть направлено на преодоление пробелов в части специальных норм об ответственности всех перечисленных специальных субъектов.

47. Особенности *административно-правовой ответственности* в рассматриваемой области определяются следующими обстоятельствами. Во-первых, использование систем ИИ может составлять часть объективной стороны «традиционных» правонарушений (как, например, в случае с нарушениями правил дорожного движения, нарушением принципов и правил обработки персональных данных, в том числе, с использованием исключительно автоматизированной обработки персональных данных и т.п.), что влечет за собой проблемы в правовой квалификации такого рода правонарушений, обусловленные технологической сложностью и субъектным составом, связанным с совершенным правонарушением. Во-вторых, в современных исследованиях подчеркивается необходимость закрепления новых специальных составов, непосредственно связанных с особенностями области ИИ. Помимо «очевидных» специальных нарушений (технологии производства, правил эксплуатации, возможной разрешительной системы), может быть необходимость и в ответственности за использование систем ИИ не по назначению, эксплуатации без специального разрешения (при условии разработки правил в отдельных областях, предусматривающих наличие такого специального разрешения). Отдельной экспертной оценки заслуживают потенциальные нарушения, связанные с ИИ в области персональных данных и антимонопольного права (в связи с автоматизированными действиями, которые приводят к злоупотреблению доминирующим положением на рынке).

48. Особенности *уголовно-правовой ответственности* также определяются тем, что системы ИИ могут рассматриваться в контексте объективной стороны преступлений. Так, например, широко известны примеры использования технологий deepfake с использованием ИИ, отдельные способы осуществления кибератак с использованием сети ботов и самообучающихся алгоритмов управления и т.п. Перечень преступлений, которые могут быть совершены с использованием систем ИИ, потенциально открытый, и уже известные системы могут быть использованы для совершения преступных посягательств потенциально на все

охраняемые уголовным законом объекты. При этом следует учитывать, что многие из уголовных правонарушений, которые могут быть совершены с использованием ИИ, уже фактически охватываются действующими составами.

Совершение гражданско-правовых сделок с использованием систем искусственного интеллекта

49. В настоящий период применение автономных систем ИИ в гражданском обороте *не осуществляется*. В этой связи можно ли говорить о потенциальных рисках их использования, а также, моделируя особенности совершения сделок с использованием систем ИИ, предлагать определенные правовые меры по формированию правовых условий заранее?

50. Гражданский кодекс РФ прямо закрепляет, что сделками признаются действия граждан и юридических лиц (ст. 153 ГК РФ). В связи с этим есть риск непризнания в качестве сделки результат взаимодействия двух или нескольких систем ИИ без участия человека, при том, что будет определена направленность на установление, изменение или прекращение гражданских прав и обязанностей в качестве сделки. Создастся неопределенность в вопросе квалификации автоматизированного взаимодействия в качестве сделки и вообще наличии договорных отношений. Кроме того, ст. 420 Гражданского кодекса РФ устанавливает, что договором признается соглашение двух или нескольких лиц об установлении, изменении или прекращении гражданских прав и обязанностей. То есть эта норма также однозначно предполагает взаимодействие лиц.

51. Представляется, что отсутствие специальных положений, устанавливающих возможность совершать сделки без участия человека, может создать риск возникновения большого количества споров об оспаривании договоров, совершенных системой ИИ. Это может стать фактором, дестимулирующим развитие ИИ. Важно отметить, что сделка обладает свойствами факта, т.е. явления объективной действительности, которое может быть воспринята другими людьми[26]. Вопрос восприятия совершения сделки связан с обеспечением прозрачности и объяснимости операций системы ИИ.

52. Возникновение наибольших рисков исследователи предполагают при эксплуатации обучаемых технологий ИИ. Главной обсуждаемой проблемой в этой области является обеспечение прозрачности принятия таких решений. При этом отмечается, что даже возможное обеспечение прозрачности имеет, тем не менее, ограниченную помощь для объяснения и понимания этих систем[27]. Одним из ограничений, связанных с прозрачностью компьютерного кода, является наблюдение за алгоритмом во времени в адаптивных системах,

[26] Скловский К.И. Сделка и ее действие. Комментарий главы 9 ГК РФ. Принцип добросовестности. 4-е изд., доп. Москва: Статут, 2019// СПС КонсультантПлюс

[27] Ananny M, Crawford K. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society. 2018;20(3):973-989. doi:10.1177/1461444816676645

которые все время обучаются и изменяются на основе полученных данных. Считается, что из-за сложности большинства алгоритмов, в частности в контексте машинного обучения, прозрачность компьютерного кода делает ее бесполезной для пользователей, стремящихся установить некоторую форму алгоритмической подотчетности. Другими словами, сделать доступными алгоритм или лежащие в его основе данные еще не позволяет считать систему подотчетной [28].

53. Анализ российской *судебной практики* в исследуемой сфере показал, что суды не квалифицируют договоры, в которых оферта формируется компьютерной программой, как автоматизированные. При этом, делая выводы о заключении или не заключении такого договора, когда оферта была размещена с существенной ошибкой, например, ценой товара в 1 руб., судьи применяют различные подходы. Так, например, в одном из дел суд признал договор незаключенным в связи с тем, что не была согласована цена товара, поскольку она была указана с ошибкой, т.е. отсутствовала воля одной из сторон на заключение такого договора[29], в другом же аналогичном деле суд презюмировал волю на заключение договора и признал его юридическую силу[30].

54. Тем не менее, суды признают ответственность пользователей программного обеспечения за их действия и ошибки. В рамках рассмотрения дел о техническом овердрафте был сделан вывод, что программные ошибки не являются следствием действия непреодолимой силы и условия договора об отсутствии вины пользователя программы в случае ее ошибки являются ничтожными. Также в отдельных случаях, в частности, в практике ФАС, должностное лицо, в обязанности которого входит исполнение определенных обязательств, если такое лицо полностью полагается на действия и решения программного обеспечения, в случае допущения им ошибок, будет нести ответственность за правонарушения в форме умысла.

55. В литературе[31] выделяют разные виды договоров (сделок)[32], которые совершаются без непосредственного участия человека: автоматизированные сделки или также используется термин «алгоритмические договоры» и «умные» договоры (smart contract). Предлагается

---

[28] Ibid.

[29] Например, решение Краснооктябрьского районного суда г. Волгограда от 19 сентября 2019 г. по делу № 2-2461/2019

[30] Например, решение Ворошиловского районного суда г. Волгограда от 11 июля 2019 г. по делу № 2-1569/2019

[31] Scholz Lauren. Algorithmic Contracts (October 1, 2016). Stanford Technology Law Review, Vol. 20, 2017// https://ssrn.com/abstract=2747701

[32] В настоящем отчете термины «сделка» и «договор» будут употребляться как равнозначные, поскольку, во-первых, анализируется в основном зарубежная литература, в которой употребляется термин «transaction», переводимый, как правило, как «сделка», и термин «contract», переводимый обычно как «договор», во-вторых для целей исследования имеет значение объединяющих их признак – направленность на установление, изменение или прекращение гражданских прав и обязанностей.

использовать общий термин для такого рода договоров – «цифровые договоры (контракты)»[33]. «Smart» контракты выделяются отдельно в этом списке цифровых контрактов, поскольку они совершаются на основе технологии распределенных реестров (блокчейн). Данные сделки подробно не изучаются в рамках данной темы, однако при описании опыта зарубежных стран они будут затронуты, поскольку исследовательская группа прогнозирует, что в будущем будут функционировать цифровые экосистемы, в которых все технологии будут взаимодействовать в соответствии со своим функциональным назначением в цифровой бизнес-среде.

В Российской Федерации в период с 1999 по 2006 год было разработано пять законопроектов «Об электронной торговле»[34], в целом соответствующих Типовому закону ЮНСИТРАЛ об электронной торговле[35], в версиях которого, начиная с 2005 года, закреплялось понятие автоматизированной сделки. В частности, предлагалось предусмотреть, что, если стороны не договорились об ином, договор может быть заключен в результате технического взаимодействия автоматизированной информационной системы и физического лица или взаимодействия автоматизированных информационных систем, при котором операции, осуществляемые такими системами, происходят вне контроля со стороны физического лица. Тем самым, главным признаком автоматизированных сделок предлагалось считать их осуществление *вне контроля со стороны физического лица.*

56. Исследование зарубежной практики показало, что опыт США может существенно использоваться для формирования правил совершения сделок с использованием систем ИИ. В США правовое регулирование автоматизированных договоров берет начало в 1999 году, когда был принят типовой закон об электронных сделках (Uniform Electronic Transactions Act; далее – UETA[36]), регулирующий отношения, возникающие в ходе деловой, коммерческой или управленческой деятельности, осуществляемой в электронной форме, за некоторыми исключениями[37]. На сегодняшний день типовой закон принят всеми штатами[38]. В 2000 году был принят типовой закон об обращении компьютерной информации (Uniform Computer

---

[33] Rizzi, Marco and Skead, Natalie, Algorithmic Contracts and the Equitable Doctrine of Undue Influence: Adapting Old Rules to a New Legal Landscape (September 23, 2020). Journal of Equity, Vol. 14, N. 3 (Forthcoming, December 2020)// https://ssrn.com/abstract=3697726

[34] Опыты цивилистического исследования: сборник статей / А.Е. Агеенко, И.И. Акимова, В.А. Волгина и др.; рук. авт. кол. и отв. ред. А.М. Ширвиндт, Н.Б. Щербаков. М.: Статут, 2018. Вып. 2.// СПС КонсультантПлюс

[35] Типовой закон ЮНСИТРАЛ об электронной торговле. Принят в г. Нью-Йорке 28.05.1996 - 14.06.1996 на 29-й сессии ЮНСИТРАЛ// доступно: СПС КонсультантПлюс

[36]Одобрен и рекомендован к принятию штатам Национальной конференцией по унификации законодательства штатов (National Conference of Commissioners on Uniform State Laws). Uniform Electronic Transactions Act 1999. § 2(6), 7A (pt. 1) U.L.A. 28. Supp. 2001. https://www.uniformlaws.org/committees/community-home/librarydocuments?communitykey=2c04b76c-2b7d-4399-977e-d5876ba7e034&tab=librarydocuments&LibraryFolderKey=&DefaultView= (дата обращения: 28.08.2020).

[37] UETA. Sec. 3(b).

[38] URL: https://www.uniformlaws.org/committees/community-home?CommunityKey=2c04b76c-2b7d-4399-977e-d5876ba7e034 (дата обращения: 28.08.2020).

Information Transactions Act; далее – UCITA[39]), который также регулирует заключение автоматизированных договоров. Оба закона, UCITA и UETA, устанавливают понятие *автоматизированного взаимодействия (automated transaction)*: взаимодействие, совершаемое или исполняемое, полностью или в части, с использованием электронных средств или записей[40], при условии, что действия или записи одной или двух сторон не пересматриваются физическими лицами в обычном порядке при заключении сделки, при ее исполнении или при исполнении обязательства, предусмотренного сделкой[41]. Типовые законы устанавливают также понятие *«электронные агенты» (electronic agents)*, под которыми понимаются компьютерная программа или электронные и иные автоматизированные средства, используемые для инициирования действия или для реагирования на электронные записи или действия без вмешательства со стороны человека[42].

57. *Классификация систем ИИ* в зависимости от степени автономности для целей регулирования сделок представляется значимой. В случае если сторона использует программу, которая действует без участия человека, но в рамках заложенного алгоритма, можно утверждать, что воля стороны на заключение договоров данной программой была выражена путем, во-первых, формирования конкретного алгоритма; во-вторых, использования программы. По сути, программа, за некоторыми исключениями и при отсутствии ошибок, не принимает каких-либо решений, которые изначально не были заложены. В то же время, когда речь идет об использовании систем ИИ, способных самостоятельно анализировать информацию, самообучаться и выходить за рамки изначально заложенного алгоритма, могут возникать проблемы, связанные с тем, что изначально запроектированные в системе ИИ параметры совершения сделок могут отличаться от итоговых решений, принимаемых этой системой.

58. Применительно *к порядку формирования воли* при совершении сделки с использованием системы ИИ на первый план выходит требование о том, что субъекты должны быть проинформированы в отношении использования систем ИИ. То есть изученный материал позволяет говорить о соблюдении информационной прозрачности отношений с использованием систем ИИ, речь идет об уведомлении лиц, что они являются участниками

[39] Одобрен и рекомендован к принятию штатам Национальной конференцией по унификации законодательства штатов (National Conference of Commissioners on Uniform State Laws). Uniform Computer Info. Transactions Act. § 107(d) 7 (pt. I) U.L.A. 9. Supp. 2001. // Режим доступа: https://www.uniformlaws.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=014d5664-8c31-a185-6e56-f1bbd7163916&forceDialog=0 (дата обращения: 28.08.2020).

[40] Под записью понимается информация, которая записана на материальном носителе или хранится на электронном или ином носителе и может быть извлечена в воспринимаемой форме, и которая создается, отправляется, передается, принимается или хранится с помощью электронных средств. UETA. Sec. 2(7)(13).

[41] UETA. Sec. 2(2). UCITA. Sec. 102(7)

[42] UETA. Sec. 2(6). UCITA. Sec.102(27).

отношений с применением системы ИИ. Представляется, что такое уведомление должно быть проектируемым и не создавать препятствий, не увеличивать сроки вхождения в договорные отношения. Риск недоверия системам ИИ в этом случае должен решаться не нормами договорного права.

59. Применительно к отношениям, в которых придется оценивать операции системы ИИ, признавать их определенными действиями по формированию воли, оценкой этого процесса будет, прежде всего, выяснение того, как запрограммирован выбор приоритетов, определены дальнейшие действия, исходя из выбранных приоритетов и, наконец, принято решение о заключении сделки и формировании соответствующих ее условий.

60. Особенность совершения сделок системой ИИ заключается в том, что в цифровой среде необходима система *непрерывной цифровой фиксации операций*, осуществляемых технологиями, всех данных, которые система использует и создает, а также соответствующего хранения всех этих данных. Также необходимо поддерживать эту систему, осуществляя периодический мониторинг на ее соответствие качеству, надежности и безопасности. Это требуется также в целях минимизации рисков оспаривания заключенных сделок, необходима проектируемая система фиксации всех операций системы ИИ, которая будет доступна для восприятия человеком в случае необходимости для представления в качестве доказательства по делу в случае спора либо контролирующему органу.

61. Все лица, которые, так или иначе, являются участниками отношений, связанных с созданием и эксплуатацией систем ИИ, применяемых для совершения сделок, должны иметь *одинаковый уровень правовой защиты* и порядок осуществления прав, независимо от того, осуществляются сделки обычным способом или в цифровой среде.

62. Широкое использование технологий ИИ в гражданском обороте может привести к ситуациям, когда их непрозрачность и широкий круг вовлеченных в эти отношения субъектов сделают чрезмерно дорогостоящим и затруднительным определение того, кто контролировал риски использования системы ИИ. Эта проблема усугубляется взаимосвязанностью цифровых технологий, их зависимостью от внешних данных и уязвимостью к нарушениям кибербезопасности. В этой связи необходимо определение сфер и *видов деятельности, относящихся к высокорисковым*, а заключение сделок с использованием систем ИИ в них осуществлять предварительно в тестовом режиме для предварительного решения вышеназванных рисков.

63. Система ИИ *представляет высокий риск, когда ее автономные действия предполагают вероятность выхода за рамки того, что можно было разумно ожидать*, а также если ее разработка, внедрение и использование влечет наступление риска причинения значительного вреда отдельным лицам или обществу, нарушения основных прав личности,

требований безопасности, учитывая сектор, в котором они разрабатываются или используются, и конкретную цель их применения.

64. Целесообразно *разграничивать* сделки, совершенные системой ИИ *самостоятельно*, и сделки, совершенные человеком *с использованием системы ИИ*. Применение на практике второй группы сделок не потребует изменений в действующее законодательство, поскольку не в полной мере автономная система ИИ будет рассматриваться как средство, используемое в договорных отношениях.

65. Необходимо будет признавать, что, когда стороны начинают использовать систему ИИ, они *выражают согласие* с условиями договоров, которые будет заключать ИИ. Волеизъявление при совершении сделок системой ИИ расширяется. Пользователи системы заранее согласны на получение результата, получаемого системой ИИ. В случае возникновения спорных ситуаций для усмотрения воли сторон должны быть доступны данные, которые позволят установить программируемые функции системы ИИ, но с учетом требований об охране интеллектуальной собственности.

66. При заключении сделок, совершаемых ИИ, обязательным требованием следует закрепить *обеспечение осведомленно*сти о заключении сделки с использованием системы ИИ. Осведомленность должна поддерживаться в любой момент договорных отношений с обеспечением быстрого выхода из них, а также предупреждением о том, что после определенных действий вернуться в начальное положение уже будет нельзя.

Алгоритмическая прозрачность в процессах принятия решений

67. Острота проблемы прозрачности алгоритмических решений *подтверждается на практике*. В частности[43], союз, представляющий интересы частных курьеров и водителей по найму в Великобритании, подал иск против Uber в окружной суд Амстердама 26 октября 2020 года, в котором оспаривается использование алгоритма для принятия решений об увольнении без консультации с человеком. Иск основывается на статье 22 GDPR, которая закрепляет, что субъект данных имеет право не подвергаться решению, основанному исключительно на автоматизированной обработке. Кроме того, данная статья предоставляет право требовать пересмотра человеком автоматизированного решения, выражать свою позицию и оспаривать это решение. В иске утверждается, что алгоритм уволил четырех водителей Uber путем направления текстового сообщения о том, что их учетные записи были деактивированы. При этом Uber не предоставил водителям ни доступа к каким-либо предполагаемым

[43] https://georgetownlawtechreview.org/in-new-european-lawsuit-uber-drivers-claim-companys-algorithm-fired-them/GLTR-11-2020/

доказательствам против них, ни возможности оспорить решение Uber о прекращении их работы.

68. Одним из проявлений повсеместной автоматизации является широкое применение автоматических или автоматизированных систем для выполнения функций, связанных с принятием решений, имеющих юридическое значение. При этом складывающаяся практика разработки и применения систем *алгоритмического принятия решений* (далее – АПР) показывает, что в процессах обработки данных, используемых для принятия юридически значимых решений, а также в самих алгоритмах установления закономерностей и выводов на их основе существуют недостатки. В частности, выборки данных, на основе которых принимаются решения, могут являться недостоверными, а сами алгоритмы принятия решений – несбалансированными, что приводит к ложным выводам и некорректным результатам. Логика, по которой система АПР приходит к тому или иному выводу, не всегда объясняется либо принципиально не объяснима. В таких условиях возникает естественный конфликт между интересами лиц, в отношении которых применяются системы АПР и которые заинтересованы в максимальной прозрачности применяемых механизмов, и разработчиками соответствующих решений, заинтересованными в снижении административного бремени при внедрении инновационных технологий.

69. Под алгоритмом, применяемым в системах АПР, понимается совокупность выраженных в коде и последовательно выполняемых инструкций, применяемых на основе данных для достижения определенной цели. В свою очередь, система АПР представляет собой совокупность алгоритмов, обеспечивающую анализ большого количества данных для установления корреляций и получения информации, являющейся полезной для принятия решения. Системы АПР включают системы автоматизации решений (*decision automation systems*), результатом деятельности которых являются готовые решения, и системы поддержки принятия решений (*decision support systems*), которые лишь оказывают человеку содействие в принятии решений[44].

70. Впервые системы АПР возникли в контексте развития экспертных систем в середине XX века как результат теоретических исследований на стыке областей организационного менеджмента и информационных технологий[45]. С тех пор рынок применения систем АПР стремительно развивался. Так, информационный онлайн-ресурс Capterra[46], нацеленный на обзор американского рынка программного обеспечения, упоминает о 88 продуктах, использующихся для поддержки принятия решений в сфере бизнеса.

[44] Springer Handbook of Automation. Editors: Nof, Shimon Y (Ed.). 2009. PP. 1574-1575.

[45] Gill T. G. Early expert systems: Where are they now? // MIS quarterly. – 1995. – P. 51.

[46] Decision Support Software // Capterra [Электронный ресурс]. – URL: https://www.capterra.com/decision-support-software/

Исследовательская компания Gartner опубликовала отчет,[47] в котором спрогнозировала, что к 2030 году системы поддержки принятия решений превзойдут по популярности все другие инициативы в области информационных технологий, станут наиболее ценными для бизнеса и будут составлять 44% рынка производства систем ИИ.

71. Алгоритмы делятся на созданные человеком и сгенерированные автоматически; детерминированные и недетерминированные; адаптивные и неадаптивные. Кроме того, системы АПР можно классифицировать в зависимости от применения в их архитектуре технологий ИИ и, в частности, машинного обучения[48]. Виды используемых алгоритмов естественным образом оказывают влияние на системы АПР, основанные на них. Дополнительно системы АПР могут быть разделены на следующие категории, в зависимости от степени участия человека в принятии решения, являющегося целью применения алгоритма[49]:

- «*человек в периметре*» (*human-in-the-loop*) предполагает, что человек сохраняет полный контроль над итоговым решением, а система только предоставляет рекомендации или входные данные. Решения не могут быть приняты без активных действий со стороны человека. Например, врач может использовать систему АПР для определения возможных диагнозов и методов лечения незнакомого заболевания. Однако окончательное решение о диагнозе и соответствующем лечении будет принимать врач. Эта модель требует, чтобы система предоставляла достаточно информации для принятия человеком обоснованного решения (в том числе, информацию о факторах, которые используются при принятии решения, их значении и удельном весе, их взаимной корреляции).

- «*человек вне периметра*» (*human-out-of-the-loop*) предполагает, что человек не контролирует принятие решений. Система АПР имеет контроль над итоговым решением без возможности человека внести какие-то корректировки.

- «*человек над периметром*» (*human-over-the-loop*) предполагает, что человеческий надзор присутствует, но лишь в определённой мере: человек имеет возможность взять на себя контроль, когда, например, модель ИИ, заложенная в систему АПР, сталкивается с неожиданными или нежелательными событиями. Такой подход позволяет человеку корректировать параметры в процессе работы алгоритма.

72. Алгоритмы могут быть сформулированы человеком или же являться результатом автоматической генерации, например, системой ИИ. При этом системы ИИ могут

[47] Gartner Says AI Augmentation Will Create $2.9 Trillion of Business Value in 2021 // Gartner [Электронный ресурс]. – URL: https://www.gartner.com/en/newsroom/press-releases/2019-08-05-gartner-says-ai-augmentation-will-create-2point9-trillion-of-business-value-in-2021

[48] Springer Handbook of Automation. PP. 1574.

[49] Model AI Framework // The Software Alliance [Электронный ресурс]. – 2019 – URL: https://ai.bsa.org/wp-content/uploads/2019/09/Model-AI-Framework-First-Edition.pdf.

использовать технологии, позволяющие системе автоматически формировать и изменять алгоритмы решения тех или иных задач посредством анализа определенных данных (так называемые технологии машинного обучения). На практике такого рода обучение может подразделяться на несколько видов, выделяемых в зависимости от степени и характера участия человека в обучении системы. Традиционно выделяются[50]:

- *контролируемое обучение* (*supervised learning*), которое предполагает формирование модели на основе отобранных (маркированных) человеком данных,

- *неконтролируемое обучение* (*unsupervised learning*), которое идентифицирует шаблоны и структуры на основе анализа немаркированных данных, а также

- *обучение с подкреплением обучения* (*reinforced learning*), предполагающее использование обучаемой машиной обратной связи о результатах её воздействия на внешнюю среду.

73. По критерию сфер применения систем АПР их также можно разделить в зависимости от юридического значения решений, принимаемых системой АПР, или с её использованием:

- системы АПР, участвующие в принятии юридически значимых решений, *непосредственно определяющие права и обязанности* того или иного лица. К данной разновидности можно отнести решения, непосредственно влекущие заключение или прекращение договоров, применение санкций и определение их размера.

- системы АПР, участвующие в принятии решений, *не имеющих непосредственного юридического значения.* К этой группе можно отнести, например, системы, формирующие необязательные рекомендации относительно назначение медицинского лечения. В юридической плоскости такие решения образуют лишь фактор, оказывающий влияние на качество (эффективность) выполнения определенного действия.

74. Требования к системам АПР делятся на общие (применяются к любой алгоритмизированной системе) и специфичные только для систем АПР. Последние, в свою очередь, могут быть классифицированы на внутренние (присущи самому алгоритму в контексте его функционирования, например, справедливость и отсутствие дискриминации) и внешние. К внешним, среди прочего, относится понятность (*understandability*), то есть возможность получить понятную информацию о связи между входными и выходными данными алгоритма. Понятность системы АПР, в свою очередь, имеет две формы: прозрачность (*transparency*) и объяснимость (*explainability*).

[50] Understanding algorithmic decision-making: Opportunities and challenges // European Parliament [Электронный ресурс]. – 2019. – URL: https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf.

75. Под *прозрачностью* системы АПР понимается доступность документации о коде программы, а также сведений о параметрах и наборе данных, используемых для её обучения (если для системы АПР характерно применение машинного обучения). В отличие от прозрачности, объяснимость требует предоставления информации не только о самой системе, но и о причинах, по которым она принимает те или иные решения. При этом объяснимость рассматривается не только как способность объяснять технические процессы. Объяснения могут касаться, например, выбора сферы применения системы АПР.

76. Объяснения относительно работы системы АПР могут принимать различную форму в зависимости от целевой аудитории, целей и контекста объяснений. В частности, объяснения могут быть операционными (информируют о том, как работает система), логическими (информируют о логических связях между входными данными и результатами) или каузальными (информируют о причинах результатов). Одновременно, в зависимости от масштаба, объяснения могут быть либо глобальными (даются в отношении всего алгоритма), либо локальными (предоставляются в отношении конкретных результатов его применения). Также объяснения могут принимать различные визуальные формы (древа решений, гистограммы, изображения или текста, примеров или примеров «от обратного» и другие).

77. Следует учитывать, что объяснимость представляет собой характеристику, которая может достигаться в разной степени. Для оценки объяснимости могут использоваться такие критерии как разборчивость, понимаемая как степень доступности объяснений для человеческого восприятия; точность; степень детализации; полнота и непротиворечивость. В то же время большинство указанных критериев носит субъективный и относительный характер, что затрудняет разработку объективных внешних инструментов оценки объяснимости.

78. Выделяется три подхода к обеспечению выполнения требования объяснимости:

- «чёрный ящик» (*black box*). Этот подход анализирует поведение системы АПР без получения информации о её коде. Объяснения строятся на основе наблюдений за отношениями между входными и выходными данными системы. Это единственно возможный подход, когда оператор или разработчик системы АПР не соглашается раскрывать её код.

- «белый ящик» (*white box*). В отличие от подхода «черного ящика», этот подход предполагает, что анализ кода системы АПР возможен.

- «конструктивный подход» (*constructive approach*). В отличие от первых двух подходов, которые предполагают, что система АПР уже существует, конструктивный подход заключается в разработке системы АПР с учетом требований объяснимости (*explainability by design*). Для достижения объяснимости как элемента дизайна возможны два варианта: опора на алгоритмическую технику, которая в силу своих характеристик отвечает требованиям

понятности, обеспечивая достаточную точность, или усиление системы АПР средствами объяснения, чтобы она могла генерировать в дополнение к своим основным результатам (например, классификации) понятное объяснение этих результатов.

79. С указанными выше характеристиками тесно связано требование подотчетности (*accountability*) – обязанность обосновать действия, совершаемые с использованием системы, и претерпевать санкции в случае неудовлетворительности обоснования. Строго говоря, прозрачность и объяснимость являются необходимыми предпосылками подотчетности, так как в отсутствие прозрачности обеспечить подотчётность практически невозможно[51].

80. В настоящее время *системы АПР широко применяются* в таких сферах как:

- здравоохранение (для помощи врачам при постановке диагноза, обследовании, назначении лекарственных препаратов, наблюдении за состоянием здоровья пациентов и автоматического оповещения медицинского персонала);

- государственное управление (для ускорения и снижения ошибок в процессе ответа на жалобы и административные заявления, принятия справедливого решения в области социального обеспечения или миграционного контроля, оповещения при чрезвычайных ситуациях, обнаружения наиболее криминогенных зон города);

- менеджмент и управление бизнес-процессами (для планирования рекламной кампании, в частности для определения выгодных уровней ценообразования и расходов на рекламу, в рамках предотвращения мошеннических действий в банковской сфере, при подборе персонала, а также для анализа активности потребителей при создании новых услуг);

- судопроизводство (для прогнозирования исхода дела, назначения наказания).

81. Применение АПР предполагает не только ряд возможных преимуществ для бизнеса и сферы государственного управления в виде сокращения временных и транзакционных издержек при принятии решений, но и ряд рисков для граждан и других участников оборота. При этом подобные риски в ряде случаев могут носить неочевидный либо отдаленный характер и в этой связи не осознаваться или не приниматься во внимание использующими АПР лицами. Кроме того, потенциал АПР, выражающийся в возможности масштабирования его применения к множеству ситуаций и лиц, создает условия для массового нарушения прав и законных интересов граждан и иных участников оборота.

82. Так, несмотря на повсеместное успешное внедрение систем АПР в сфере здравоохранения, известны несколько случаев обнаружения ошибок. Так, база данных для прогнозирования рака молочной железы, содержащая в большей мере снимки белокожих

[51] A governance framework for algorithmic accountability and transparency // European Parliament [Электронный ресурс]. – 2019. – URL: https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf.

женщин, не учитывала различные факторы, релевантные для темнокожих женщин, а также тот факт, что темнокожие женщины склонны к раку молочной железы на 42% чаще[52]. Это пример как нерепрезентативной выборки, так и выборки, содержащей большое количество несущественных признаков при применении модели к темнокожим женщинам. Другой пример – обнаружение неточностей в системе для скрининга злокачественных опухолей[53]. Из-за того, что на изображениях в базе данных раковых больных зачастую присутствует линейка, при помощи которой измерялся размер опухоли, алгоритмы ошибочно стали классифицировать любое изображение с линейкой как злокачественную опухоль.

83. Отдельного упоминания в контексте рисков применения систем АПР заслуживает проект COMPAS, предназначенный для прогнозирования рецидива у подсудимых в США. В отличие от электронного правосудия, при котором судебное заседание переносится в формат видеоконференции, предикативное правосудие нацелено i) на полное исключение из судебного процесса судьи-человека, заменив его программным обеспечением, разрешающим дело на основе специфических параметров и информации, полученной из корпуса предыдущих судебных решений внутри заданной категории дел, или ii) на частичное делегирование судьей рабочих практик системе. Ко второму типу предикативных систем и относится проект COMPAS, приложение которого использует общедоступные статистические данные, а также данные заполненного обвиняемым опросника из 137 вопросов, по результатам анализа которых выставляется скоринговая оценка лица, определяющая вероятность рецидива. Каждый обвиняемый до суда получает не менее трех значений рейтинга: «риск рецидива», «риск насилия» и «риск неявки». Баллы по системе COMPAS для каждого обвиняемого варьируются от 1 до 10: баллы от 1 до 4 помечаются COMPAS как низкий риск, от 5 до 7 – как средний риск, от 8 до 10 баллов – как высокий риск. Баллы означают вероятность повторного совершения лицом правонарушения, что рассчитывается на основании сравнения данных об обвиняемом со средними данными об осужденных по соответствующим статьям.

84. Журналистское расследование компании ProPublica, проведенное в 2016 году на выборке из более 10 000 обвиняемых, обнаружило[54], что многие чернокожие обвиняемые были ошибочно отнесены к группе высокого риска, а белые обвиняемые чаще, чем чернокожие подсудимые, категорируются как лица с низким уровнем риска. ProPublica сравнила результаты системы COMPAS с показателями рецидивизма обвиняемых в течение

[52] Google's AI for mammograms doesn't account for racial differences // Quartz [Электронный ресурс]. – URL: https://qz.com/1781123/googles-ai-for-mammograms-doesnt-account-for-race/

[53] Kelly C. J. et al. Key challenges for delivering clinical impact with artificial intelligence // BMC medicine. – 2019. – Vol. 17. – №. 1. – PP. 1-9.

[54] How We Analyzed the COMPAS Recidivism Algorithm // ProPublica [Электронный ресурс]. – URL: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

двух лет и выявила, что система правильно предсказывала рецидивизм только в 61% случаев, а по такой категории как насильственный рецидив – только в 20% случаев.

85. В целом, в качестве *основных рисков применения систем АПР* можно выделить:

- риски для неприкосновенности частной жизни (функционирование систем ИИ, лежащих в основе большинства систем АПР, требует сбора и обработки значительных массивов данных, которые могут содержать персональные данные граждан, в том числе и в обезличенном виде);

- риски дискриминации граждан при использовании АПР (дискриминационные начала могут появиться в алгоритмах обработки данных в силу двух основных причин: наличия предубеждений у разработчиков алгоритма, которые они намеренно или бессознательно перенесли в его архитектуру, а также недостатки качества входных данных);

- риски, связанные с нарушением права на справедливый суд;

- риски манипулирования индивидуальным и общественным мнением (алгоритмическая обработка данных и принятые на основе нее решения могут не только ограничивать граждан в самостоятельном формировании своих предпочтений, но и подталкивать их к совершению определенных действий, как в коммерческом плане, так и в публично-правовой сфере);

- риски принятия неправильных (неэффективных) управленческих решений;

- риски, связанные с привлечением к ответственности за результаты применения технологий АПР.

86. Подавляющее большинство упомянутых рисков могут быть если не устранены, то существенно минимизированы за счет привнесения прозрачности и объяснимости в процессы разработки и эксплуатации систем АПР. Как отмечается, прозрачность может исправить ошибки в алгоритмических процессах, способствуя тем самым их эффективности. Она позволяет лицам исправлять неточные данные, а также вынуждать операторов алгоритмов улучшать свои практики. Наличие знания о причинах вынесенного дискриминационного решения (проблемы в качестве исходных данных либо качестве самого алгоритма) способно существенным образом расширить возможности по оспариванию такого решения и защите прав потерпевшего от него субъекта. Такого рода механизмы прозрачности могут быть реализованы в различных формах: как на уровне жесткого права, так и в виде рекомендаций и иных подобных инструментов «мягкого права». При этом следует особо подчеркнуть, что возможные инициативы по саморегулированию или выработке «мягкого права» в данной области не должны подменять необходимость выработки адекватного нормативного регулирования в этой сфере, поскольку вопросы защиты прав человека не могут быть делегированы на уровень саморегулирования.

87. Существующее в Российской Федерации законодательное регулирование вопросов использования систем АПР находится *на начальных этапах* своего формирования. В этой связи проанализированный зарубежный опыт разработки специализированных правовых норм, а также связанная с применением АПР правоприменительная практика, приобретают особую актуальность, позволяя не только избежать ненужного дублирования усилий по подготовке законодательных предложений, но и обеспечить совместимость вырабатываемого в России регулирования с зарубежным. При этом практически полное, за редким исключением, отсутствие в России специализированных правовых норм, посвященных регулированию АПР, может даже рассматриваться как определенное преимущество, так как создает благоприятные условия для формирования нового регулирования с минимальными рисками создания коллизий с уже существующими законодательными положениями.

88. Среди мер по обеспечению прозрачности функционирования систем АПР, которые имеет смысл реализовать на уровне федерального законодательства Российской Федерации (в том числе, с целью обеспечения большего доверия субъектов воздействия систем АПР к их использованию в публичной сфере), следует выделить следующие:

- установление на нормативном уровне *классификационных критериев*, позволяющих осуществить категорирование систем АПР с учетом характерного для их применения уровня риска, с одновременным установлением дифференцированных требований, обеспечивающих прозрачность, объяснимость и подотчетность таких систем. Так, для систем высокой степени риска может быть установлено требование об обязательном логировании всех действий, происходящих в системе АПР в ходе принятия решения, в то время как для систем, сопряжённых с низкими рисками, такое требование представляется избыточным. При этом важно избежать дублирования со смежными сферами регулирования (в частности, законодательством о критической информационной инфраструктуре).

- введение требования об *обязательном осуществлении предварительной самостоятельной и/или внешней оценки алгоритмического воздействия* системы АПР на предмет справедливости, отсутствия дискриминации, точности и релевантности обучающих данных. Вид оценки, применимый к системе АПР, должен зависеть от типа рисков, характерных для системы АПР (с учетом классификационных критериев, упомянутых выше). Несложно заметить, что большинство примеров наиболее рискованного применения систем АПР относятся именно к сфере осуществления публичных полномочий. С учетом вышеизложенного обязательность оценки систем АПР для государственных органов должна быть введена в первую очередь. Экстраполяция аналогичного требования на частный сектор целесообразна после апробирования регулирования на сфере государственного управления.

- установление *запрета на использование в наиболее рискованных сферах* алгоритмов, не обеспечивающих достаточный уровень объяснимости. В частности, в системах АПР, используемых для принятия юридически значимых решений, имеющих существенное влияние на права субъектов и функционирующих без участия человека, следует ограничить использование самообучающихся алгоритмов, которые могут самостоятельно пересматривать применяемые правила.

- формирование государственными органами *регулярных отчетов* об использовании систем АПР. Как известно, наличие качественной обратной связи является критичным для обеспечения эффективного функционирования любой системы АПР. В качестве основы можно использовать опыт штата Нью-Йорк, где предлагается установить обязанность государственного органа по предоставлению ежегодных отчетов об используемых массивах данных для систем АПР. В таких отчетах должны содержаться описание используемых массивов данных, источники и способы его формирования, сведения о частоте обновлений, перечень государственных органов, имеющих к нему доступ, размеры массива данных и иные сведения. Представляется, что отчет вполне может быть расширен и за рамки собственно предоставления информации об используемых массивах данных для функционирования системы АПР, за счет включения в него сведений о полученной «обратной связи» в виде жалоб от заинтересованных лиц. Подобного рода отчеты должны иметь публичный характер, что будет соответствовать принципу открытости информации о деятельности государственных органов и свободе доступа к ней.

- *создание единого реестра систем АПР, используемых в государственном управлении*, а также *наиболее опасных* частных систем АПР. В отсутствие такого реестра формирование единой политики в отношении использования систем АПР будет существенно затруднено. Российское законодательство имеет достаточно богатый опыт создания реестров как одного из инструментов определения объектов специального регулирования, который может быть органично распространен на сферу применения АПР.

- *обеспечение аудируемости алгоритмов* используемых систем АПР, что потребует внесения специальных положений в нормы ГК РФ, которые регламентируют возможность использования охраняемых объектов авторского права (программного кода, в котором находят свое отражение алгоритмы) без согласия правообладателя и без выплаты ему вознаграждения. Такие положения, согласно существующей в России парадигме регулирования, должны быть сформулированы достаточно конкретно и узко, чтобы предотвратить злоупотребление ими конкурентами коммерческих разработчиков систем АПР.

- введение требования *об обязательном размещении уведомления* о применении системы АПР на соответствующем сайте государственного органа или иной организации,

использующей систему АПР. Одновременно желательно закрепить правило, в соответствии с которым любое решение, принятое на основе исключительно алгоритмической обработки, должно содержать отметку о том, что оно было принято без участия человека. Дополнительно могут быть установлены требования к содержанию соответствующего уведомления.

Устранение барьеров для инвестиционных проектов в сфере ИИ и робототехники

89. Исходя из существующего регулирования можно определить следующие общие формы реализации инвестиционных проектов в РФ: концессионное соглашение[55], соглашение о государственно-частном партнерстве (далее – ГЧП)[56], контракт жизненного цикла[57], инвестиционное соглашение[58], соглашение о защите и поощрении капиталовложений[59]. Вместе с тем, принимая во внимание, что базовое правовое регулирование указанных выше форм реализации инвестиционных проектов не всегда отвечает потребностям и особенностям инвестиционных проектов в области информационных технологий (ИТ), в частности в сфере искусственного интеллекта (ИИ) и робототехники, законодателем разрабатываются и принимаются новеллы в соответствующие федеральные законы.

90. Классификация правовых проблем инвестиционных проектов в сфере искусственного интеллекта и робототехники выглядит следующим образом:

- *Правовой статус искусственного интеллекта и роботов как объектов и субъектов гражданских прав.* В отечественной системе права логичным выглядит вывод о том, что искусственный интеллект и роботы являются объектами правоотношений и не выступают субъектами права, поскольку право регулирует общественные отношения между людьми как биологическими существами. В классической теории права достаточно критично относятся к идее, что искусственный интеллект может обладать признаками правосубъектности.

- *Конструкция инвестиционных договоров.* Очевидно, что инвестиционная деятельность, являясь экономическим отношением, имеет собственную правовую оболочку. Такой оболочкой, преимущественно, является гражданское законодательство. При этом при анализе сущности искусственного интеллекта и особенностей робототехники обоснованным представляется тезис, что общественные отношения в области создания, владения,

[55] См. Федеральный закон «О концессионных соглашениях» от 21.07.2005 № 115-ФЗ // СЗ РФ от 25.07.2005. № 30 (часть II). Ст. 3126.
[56] См. Федеральный закон «О государственно-частном партнерстве, муниципально-частном партнерстве в Российской Федерации и внесении изменений в отдельные законодательные акты Российской Федерации» от 13.07.2015 № 224-ФЗ // СЗ РФ от 20.07.2015. № 29 (часть I). Ст. 4350.
[57] См. Федеральный закон «О контрактной системе в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд» от 05.04.2013 № 44-ФЗ // СЗ РФ от 08.04.2013. № 14. Ст. 1652.
[58] См. Федеральный закон «Об инвестиционной деятельности в Российской Федерации, осуществляемой в форме капитальных вложений» от 25.02.1999 № 39-ФЗ // СЗ РФ от 01.03.1999. № 9. Ст. 1096.
[59] См. Федеральный закон от 01.04.2020 № 69-ФЗ «О защите и поощрении капиталовложений в Российской Федерации» // СЗ РФ от 06.04.2020. № 14 (часть I). Ст. 1999.

пользования и распоряжения искусственным интеллектом, а также производство и эксплуатация роботов, носят информационный характер. Информация, как базовая, сквозная категория информационного права, составляет «ядро» искусственного интеллекта. Для робототехники информация является тем инструментом, который позволяет отграничить робота от манекена с помощью специального программного обеспечения и базы данных. Присутствие искусственного интеллекта и его производных в экономическом обороте, с одной стороны, регламентируется нормами гражданского законодательства с точки зрения принципов, общих правил и регламентов. Однако информационная сущность искусственного интеллекта не учитывается в полном объеме, а «размывается» в инвестиционных нормах ГК РФ. Это явление приводит к тому, что в договорах инвестиционного характера достаточно проблематично надлежащим образом сформулировать предмет.

- *Защита прав иностранных инвесторов в области искусственного интеллекта и робототехники.* Во многих странах технологии искусственного интеллекта и робототехники развиваются соразмерно с «традиционными» высокотехнологичными производствами на основе принципа межгосударственного сотрудничества. Субъектный состав правоотношений в области инвестиций в искусственный интеллект и робототехнику подпадает под правовое регулирование юрисдикций многих государств.

- *Распределение рисков в инвестиционных проектах в области искусственного интеллекта и робототехники.*[60]

- *Актуализация инвестиционного законодательства.* Общими недостатками российского законодательства, влияющими на инвестиционный климат в целом, являются его бессистемность, избыточность, противоречивость, наличие многочисленных пробелов регулирования, отставание внесения изменений и приведения в соответствие с иными актами. Российское инвестиционное законодательство не приспособлено под новые экономические тренды, связанные с искусственным интеллектом и робототехникой, отсутствует какие-либо нормативные правовые акты, определяющие специфику искусственного интеллекта и робототехники как объекта инвестиционной деятельности.

- *Наличие публично-правовых ограничений в данной области.* В первую очередь речь идёт об ограничениях для иностранных инвесторов.

91. Для устранения выделенных барьеров мы могут быть предложены следующие шаги:

- *определение правового статуса ИИ и специфики правового регулирования, связанного с его использованием.* В настоящее время активы, созданные ИИ, действующим автономно,

[60] Вопрос более подробно рассмотрен в предыдущих разделах настоящего доклада.

выпадают из сферы правового регулирования и не являются охраноспособными, что может препятствовать привлечению инвестиций в данную сферу.

- *создание специальных правовых режимов для разработки новых технологий.* Введение экспериментальных правовых режимов может способствовать определению наиболее сбалансированного регулирования тех или иных отношений. По завершении эксперимента проводится оценка влияния изъятий из регулирования. Такой подход помогает в быстрые строки и с минимальными рисками для прав и свобод человека, безопасности государства и общества протестировать технологии.

- *создание специальных правовых режимов для привлечения инвестиций.* Речь идет в первую очередь о налоговых льготах. Будучи одним из основных инструментов регулирования экономики, налоговые льготы влияют на макроэкономическую обстановку в стране и уровень экономического роста. Они призваны стимулировать развитие производства во многих отраслях экономики[61]. Уменьшение налогового бремени инвесторов по спецпроектам в части налога на прибыль и добавочную стоимость не менее оправдано. Новые технологии – всегда риск. Готовность инвестировать в то, что еще не создано без четких гарантий результата, увеличится за счет минимизации издержек. Кроме того, в сфере искусственного интеллекта нередко встречается сотрудничество между компанией-заказчиком, которая готова инвестировать в технологию, но хочет провести пилот, и ИТ-стартапом. Заказчик предоставляет свои данные разработчикам и оценивает результаты апробации решения на этих данных. Участники, как правило, оформляют это соглашением о сотрудничестве. Однако формальная безвозмездность таких соглашений нередко ведет к налоговым рискам и последующему доначислению НДС и других налогов.

- *ограничение ответственности для участников инвестиционных проектов.* Риски, связанные с неопределенностью результата инвестиционных проектов в сфере искусственного интеллекта и робототехники, снижают объём финансирования. Ограничение ответственности в данной области поможет устранить барьеры, связанные с неготовностью участников инвестиционных проектов вкладывать средства и ресурсы в разработку алгоритмов, конечный результат работы которых нельзя однозначно гарантировать.

- *расширение прав на результаты инвестиционных проектов для частных партнеров.*

Устранение неопределенности относительно прав использования частным партнером результата инвестиционного проекта уберет барьеры, связанные с неготовностью разработчиков участвовать в соглашениях с государственными заказчиками.

[61] Нестеренко Ю. Н. Налоговые льготы: новые подходы к установлению // Экономический журнал. 2017. [Электронный ресурс]. URL: https://cyberleninka.ru/article/n/nalogovye-lgoty-novye-podhody-k-ustanovleniyu

- *назначение единого ответственного органа и центра управления.* Активность по всем направлениям, связанным с внесением изменений в законодательство по данной тематике, должна идти в комплексе и координироваться одним оператором. Дополнительно, через консолидирование полномочий формируется единое «окно» обращений для участников рынка.

- *создание центра сотрудничества между государством и научным и бизнес-сообществами.*

- *постановка KPI (ключевых показателей эффективности) по финансированию ИИ и робототехники для государственных органов и контроль за их исполнением.*

- *расширение доступа к данным,* что является ключевым условием для развития технологий искусственного интеллекта. Расширение доступа к данным устранит барьеры для их использования разработчиками. Юрисдикции с благоприятным правовым регулированием становятся местом концентрации технологических компаний, и, как следствие, инвесторов.

- *развитие образования в области искусственного интеллекта и робототехники.*

- *создание консультационных центров.* Барьеры правового регулирования связаны не только с наличием или отсутствием тех или иных правовых норм, но и практикой их применения. Для субъектов малого предпринимательства, которыми являются большое количество ИТ-стартапов, участие в конкурсе на получение финансирования или гранта – проблема с точки зрения прохождения конкурсных процедур. Неопределенность в отношении правомерности использования тех или иных данных для создания решений, необходимость регистрации результатов интеллектуальной собственности в подтверждение исполнения договора[62]– это лишь несколько вопросов, с которыми сталкиваются разработчики. Создание консультационных центров на базе уполномоченного органа или специального координационного центра поможет компаниям разобраться в сложных и неоднозначных вопросах и участвовать в конкурсах.

92. Подводя итог вышесказанному, исследователи Института права цифровой среды выражают уверенность, что не только развитие технологий ИИ ставит перед юридической наукой целый ряд безусловно интересных с теоретической точки зрения вопросов, но и грамотное правовое обеспечение в сфере регулирования ИИ и робототехники, учитывающее российскую юридическую традицию и зарубежный опыт, способно обеспечить лидирующую роль нашей страны в области разработки и эффективного применения наиболее перспективных технологических решений.

[62] Основные условия программы «Старт» Фонда содействия развитию малых форм предприятий в научно-технической сфере (Фонд содействия инновациям): http://fasie.ru/programs/programma-start/