ПРАВОВОЙ ФОРМАТ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Карцхия А.А.[1]

Ключевые слова: *нейронные сети, технологический суверенитет, прорывные технологии, правосубъектность, устойчивое развитие, риск-ориентированный метод.*

Аннотация

Цель исследования: анализ правовых аспектов регулирования систем и технологий искусственного интеллекта (ИИ), новых моделей ИИ для достижения целей устойчивого развития, выявления особенностей применения традиционных правовых категорий в сфере ИИ.

Методы исследования: сравнительно-правовой метод и методы анализа и синтеза в процессе исторического генезиса ИИ, применение риск-ориентированного метода оценки ИИ.

Результаты исследования: автор пришел к выводу о том, что ИИ способен преобразовать практически все аспекты экономики и социальных отношений в обществе, а его вновь открывающиеся возможности носят трансформационный и глобальный характер. Вместе с тем неординарные возможности технологий ИИ сопряжены с рисками, которые могут угрожать глобальной стабильности и подрывать общечеловеческие ценности. Для преодоления возможных угроз и рисков, смягчения потенциальных опасностей крайне востребована стратегия разработки системных правовых мер и способов регулирования технологий и моделей ИИ в национальном и международном масштабе.

Новизна исследования: на основе риск-ориентированного подхода рассматривается правовой формат эффективности, устойчивости и безопасности технологий и моделей ИИ, определение его правового статуса, в том числе с точки зрения защиты человека от неконтролируемого влияния ИИ и неизменности гарантий прав и свобод человека.

DOI: 10.21681/2226-0692-2024-2-125-133

Искусственный интеллект… взломал операционную систему нашей цивилизации.
(Юваль Н. Харари,
газета «Экономист», 28 апреля 2023 г.)

Введение

Искусственный интеллект (ИИ) находится в центре усиливающегося технологического соревнования государств мира. Как отмечается в Концепции внешней политики Российской Федерации[2], человечество переживает эпоху революционных перемен, которая связана, прежде всего, со структурной перестройкой мировой экономики, обусловленной переходом на новую технологическую основу посредством внедрения технологий ИИ, новейших информационно-коммуникационных, энергетических, биологических технологий и нанотехнологий, а также ростом национального самосознания, культурно-цивилизационного разнообразия и другими объективными факторами, которые ускоряют процессы перераспределения потенциала развития в пользу новых центров экономического роста и геополитического влияния, способствуют демократизации международных отношений.

Вслед за США и Китаем в гонку по развитию ИИ включились Великобритания, Франция, Германия, Индия, Саудовская Аравия, Объединенные Арабские Эмираты, пообещав выделить по несколько десятков миллиардов долларов на инвестиции в ИИ (см. табл. 1). Стратегии этих стран по развитию ИИ основаны не только на финансовой поддержке государства. Так, правительство США решает задачу по уменьшению зависимости Америки от тайваньских производителей полупроводников, одновременно начав процесс перевода основных производственных мощностей

[2] Указ Президента РФ от 31.03.2023 № 229 «Об утверждении Концепции внешней политики Российской Федерации» // Собрание законодательства РФ, 03.04.2023, № 14, ст. 2406.

[1] ***Карцхия Александр Амиранович,*** *доктор юридических наук, профессор РГУ нефти и газа (НИУ) имени И.М. Губкина, г. Москва, Российская Федерация.*
E-mail: arhz50@mail.ru

по производству чипов на территорию США. Кроме того, администрация президента Джо Байдена ввела жестокий экспортный контроль, который запрещает продажу передовых технологий ИИ, включая чипы и оборудование для их производства, другим государствам – прежде всего КНР и России. Этот запрет распространяется и на передачу современных технологий и ноу-хау. К примеру, ограничиваются продажи чипов для ИИ американской компании NVidia, которая является их основным производителем электроники.

Стратегия Китая в области ИИ во многом является ответом на американскую политику сдерживания. В период с 2021 по 2022 год КНР потратило почти 300 млрд долларов США на воссоздание цепочки поставок и производства чипов для ИИ и других полупроводников в национальной промышленности, что помогло, например, Huawei и Smic, крупнейшему китайскому производителю микросхем, разработать и изготовить сложный графический процессор. Осуществляются существенные государственные инвестиции в стратегически важные технологии для достижения технологических приоритетов. Правительство КНР также способствует биржевой торговле данными, где предприятия могут обмениваться коммерческими данными обо всем, от продаж до производства, что позволяет и небольшим фирмам, специализирующимся в области ИИ конкурировать с крупными фирмами. В Китае уже существует 50 таких бирж.

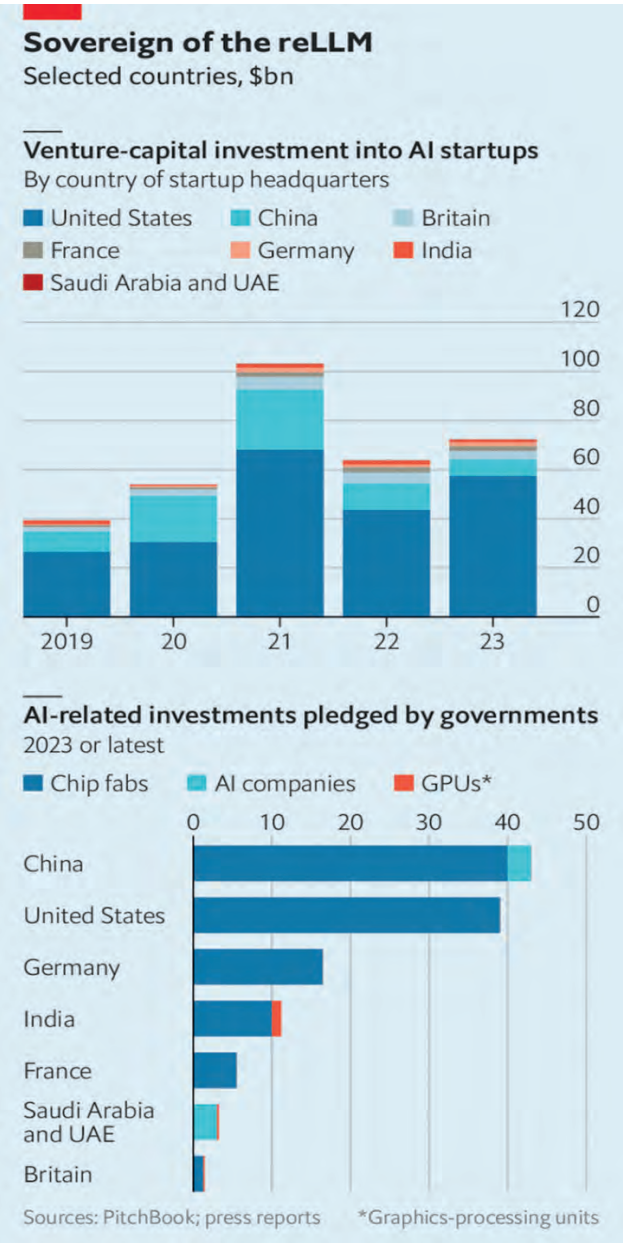
В то же время технологии ИИ занимают одно из центральных мест в структуре Глобального цифрового договора (ГЦД, англ. *Global Digital Compact)*[3], который инициативно продвигает генеральный секретарь ООН А. Гутерриш для подписания на Саммите будущего в сентябре 2024 года в рамках технологического направления с участием всех заинтересованных сторон, включая ООН и его государства-члены, частные компании, организации гражданского общества. Ожидается, что ГЦД «обозначит общие принципы открытого, свободного и безопасного цифрового будущего для всех», включая такие вопросы, как цифровое подключение, недопущение фрагментации Интернета, предоставление пользователям вариантов использования их данных, реализация прав человека в Интернете и продвижение надежного Интернета путем установления критериев ответственности за дискриминацию и вводящий в заблуждение контент.

Как отмечают эксперты [1], ГЦД продолжает практику глобальных договоров ООН, которые не являются договорами в юридическом смысле данного термина, а представляют собой набор общих принципов, призванных регулировать деятельность различных акторов мировой политики. Важной целью ГЦД видится возможность регламентации деятельности крупных ИТ-компаний, которые сегодня обладают практически неограниченной властью в глобальном цифровом обществе и, как следствие, нуждаются в международно согласованных рамках, ограничивающих возможности их злоупотребления своими ресурсами в цифровой среде. С инициативой выработать правила поведения для крупных цифровых платформ уже выступал МИД России, однако, в отличие от подхода ГЦД, Россия исходит из необходимости уважения государственного суверенитета, поэтому практическая выработка соответствующих правил и их имплементация – прерогатива государств.

Однако главным изъяном ГЦД является попытка уравнять государства с неправительственными организациями в регулировании информационно-коммуникационного пространства, в чём проявляется стремление закрепить доминирование западных IT-корпораций в ущерб интересам Глобального Юга [2].

Табл. 1

Инвестиции государств в искусственный интеллект



Источник: The Economist

[3] URL: https://www.un.org/techenvoy/ru/global-digital-compact

Глобальная конкуренция в сфере искусственного интеллекта

Новейшие технологии играют ключевую роль в развитии экономики, социальных проектов, обороны и национальной безопасности государства.

В частности, Национальная стратегия развития искусственного интеллекта РФ на период до 2030 года (в ред. 2024 года)[4] провозглашает целями развития ИИ в Российской Федерации обеспечение роста благосостояния и качества жизни ее населения, обеспечение национальной безопасности и правопорядка, достижение устойчивой конкурентоспособности российской экономики, в том числе лидирующих позиций в мире в области ИИ. Стратегия определяет ИИ как комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их. Этот комплекс включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в т. ч. такое, в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений.

Вместе с тем в Стратегии национальной безопасности Российской Федерации[5] указано, что для решения поставленных задач в сфере национальной безопасности ИИ используется как инструмент обеспечения информационной безопасности на основе применения передовых технологий, включая технологии ИИ и квантовые вычисления, как средство модернизации промышленных предприятий и инфраструктуры, цифровизации в целях повышения производительности труда; в целях научно-технологического развития России установлено также развитие таких перспективных высоких технологий, как ИИ, наряду со следующими: нанотехнологии, робототехника, медицинские, биологические, генная инженерия, информационно-коммуникационные, квантовые, обработка больших данных, энергетические, лазерные, аддитивные, создания новых материалов, когнитивные, природоподобные технологии.

Международное значение технологий ИИ еще раз отмечено Генеральной Ассамблеей ООН 21 марта 2024 г. в Резолюции по искусственному интеллекту «Использование возможностей безопасных и заслуживающих доверия систем искусственного интеллекта для устойчивого развития»[6], которая поддержана более чем 120 государствами-членами и направлена на поощрение государств мира к защите прав человека и его персональных данных при использовании технологий ИИ в контексте возможных рисков. Резолюция не имеет обязательной силы, но Устав ООН наделяет Генеральную Ассамблею полномочиями инициировать исследования и давать рекомендации по содействию в развитии и кодификации международного права. В содержании Резолюции отмечается, что безопасные, защищенные и надежные системы ИИ – под которыми понимаются системы ИИ в невоенной сфере, жизненный цикл которых включает этапы от предварительного проектирования до эксплуатации и вывода из эксплуатации, которые являются антропоцентричными, надежными, объяснимыми, этичными, инклюзивными, в полной мере уважающими, поощряющими и защищающими права человека и международное право, обеспечивающими конфиденциальность, ориентированными на устойчивое развитие и ответственными, – именно эти системы ИИ способны ускорить и стимулировать прогресс в достижении всех 17 целей в области устойчивого развития во всех его трех – экономическом, социальном и экологическом – компонентах на сбалансированной и комплексной основе.

Резолюция также констатирует, что ненадлежащее или злонамеренное проектирование, разработка, внедрение и использование систем ИИ (без соблюдения надлежащих мер безопасности или в нарушение норм международного права) создает риски, которые могут помешать прогрессу в осуществлении Повестки дня в области устойчивого развития на период до 2030 года и достижении поставленных в ней целей в области устойчивого развития и подорвать устойчивое развитие во всех его трех – экономическом, социальном и экологическом – компонентах; расширить цифровой разрыв между странами и внутри стран; усилить структурное неравенство и предвзятость; привести к дискриминации; подорвать целостность информации и доступ к ней; отрицательно сказаться на защите, продвижении и осуществлении прав человека и основных свобод, включая право не подвергаться незаконному или произвольному вмешательству в частную жизнь; увеличить потенциальный риск аварий и комплексных угроз со стороны злоумышленников.

Правовые аспекты глобального регулирования ИИ стали предметом обсуждения группы стран G7, когда 30 октября 2023 г. в г. Хиросима (Япония) они приняли совместную Декларацию *"G7 Leaders' Statement on the Hiroshima AI Process"*[7], в составе которой приняты два основных документа: свод Международных руководящих принципов по ИИ (*The International Guiding Principles on Artificial Intelligence)* и Кодекс поведения для разработчиков ИИ (*Code of Conduct for AI Developers*)[8], содержа-

[4] Утв. Указом Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» // Собрание законодательства РФ, 14.10.2019, № 41, ст. 5700.

[5] Указ Президента РФ от 02.07.2021 № 400 «О Стратегии национальной безопасности Российской Федерации» // Собрание законодательства РФ, 05.07.2021, № 27 (часть II), ст. 5351.

[6] United Nations Resolution on AI "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development", March 21, 2024. URL: https://undocs.org/

[7] URL: https://digital-strategy.ec.europa.eu/en/library/g7-leaders-statement-hiroshima-ai-process

[8] G7 HIROSHIMA PROCESS ON GENERATIVE ARTIFICIAL INTELLIGENCE (AI): TOWARDS A G7 COMMON UNDERSTANDING ON GENERATIVE AI, September 2023, OECD 2023. URL: http://www.oecd.org/termsandconditions

щий набор правил, которым рекомендуется следовать разработчикам ИИ на добровольной основе для снижения рисков на протяжении всего жизненного цикла ИИ в целях разработки безопасных и заслуживающих доверия систем ИИ, в частности, рисков от генеративного ИИ. *Основными пятью рисками признаются*: распространение дезинформации и манипулирование, нарушения интеллектуальной собственности, угрозы конфиденциальности, дискриминация и предвзятость, а также риски для безопасности. Требуется управление искусственным интеллектом на основе предложенных *Международных руководящих принципов* и *Кодекса поведения для организаций, разрабатывающих передовые системы ИИ*.

Предполагается, что свод Международных руководящих принципов по искусственному интеллекту и Кодекс поведения разработчиков ИИ будут постоянно пересматриваться и обновляться, чтобы гарантировать их актуальность, учитывая стремительный характер развития технологий ИИ. В Кодексе поведения отмечается, что различные страны могут применять в своей юрисдикции уникальные подходы к реализации правил по-своему. К примеру, для государств Евросоюза такой основой может стать Закон об искусственном интеллекте (*Artificial Intelligence Act*)[9].

В Меморандуме OECD 2023 г. система ИИ определяется как компьютерная система, которая для достижения явных или неявных целей на основе получаемых входных данных определяет, как генерировать выходные данные, такие как прогнозы, контент, рекомендации или решения, которые могут влиять на физическую или виртуальную среду. Различные системы ИИ различаются по уровню автономии и адаптивности после развертывания.

Следует отметить, что термин «искусственный интеллект» – общий, объединяющий множество технологий использования математико-статистических методов моделирования когнитивных способностей. Технологии ИИ работают на основе анализа большого объема неструктурированных данных *(Big Data)* по специально разработанному алгоритму для выявления определенных закономерностей данных и получения на их основе конкретного вывода с использованием нейронной сети, алгоритмы и структура которых основаны на функциональных принципах человеческого мозга, где большое количество отдельных алгоритмов работают вместе взаимосвязанным и взаимозависимым образом, отражающим функционирование сети синапсов в человеческом мозге. Сложные нейронные сети с несколькими уровнями обработки (со множеством соединенных последовательно и влияющих друг на друга алгоритмов) называются глубокими нейронными сетями *(Deep Neural Networks)*. В сложных («глубоких») нейронных сетях способ взаимодействия отдельных алгоритмов друг с другом больше не определяется разработчиком, поскольку количество определяемых параметров слишком велико. Вместо этого подходящие обучающие данные (т. е. обучающие данные, специально отобранные и предназначенные для использования по назначению) передаются в нейронную сеть для обработки в автоматических циклах обучения. Нейронная сеть использует процессы статистической оптимизации для определения наиболее подходящих настроек (параметризация), например, для автономной идентификации лица на снимках. Этот процесс автоматической параметризации нейронной сети известен как глубокое обучение (*Deep Learning*) [3].

В последнее десятилетие во многих государствах мира идет активная разработка правовых актов в сфере регулирования ИИ.

Так, в США 30 октября 2023 г. президент Байден издал Указ о безопасном и заслуживающим доверия ИИ *(Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence)*[10], который устанавливает новые стандарты безопасности ИИ, предусматривает комплекс практических мер и поручений госорганам по реализации конкретной политики для решения проблемных вопросов в сфере национальной безопасности, защиты данных, трудовых отношений и социального здравоохранения. Указ предусматривает обязанность компаний-разработчиков самых мощных систем ИИ сообщать результаты испытаний по безопасности ИИ и другую важную информацию правительству США. В соответствии с Законом об оборонном производстве Указ требует от компаний-разработчиков базовых моделей ИИ, потенциально представляющую серьезную угрозу национальным безопасности, экономической безопасности или общественному здравоохранению, уведомлять федеральное правительство при обучении модели ИИ о результатах всех пентестов (*red team*) на оценку кибербезопасности модели ИИ до того, как компании обнародуют эти результаты. Этот указ издает сотни директив, касающихся безопасности ИИ, более чем двадцати федеральным агентствам, ставя перед ними задачи по реализации конкретной политики для решения проблемных областей, таких как национальная безопасность, защита данных, предвзятость на рабочем месте и общественное здравоохранение. Он также налагает требования на частные компании, разрабатывающие мощные системы ИИ, которые могут представлять угрозу национальной безопасности или общественному здоровью, требуя от них делиться результатами и методами испытаний на безопасность и другой важной информацией с правительством США. Большинство директив, изданных президентом Байденом в соответствии с Исполнительным указом, должны быть выполнены в течение следующего года.

В марте 2024 г. Европарламент принял Закон ЕС об искусственном интеллекте (*EU Artificial Intelligence*

[9] URL: https://artificialintelligenceact.eu/

[10] URL: https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

Act)[11] (далее – Закон), который будет сочетаться с существующими правовыми нормами других законов ЕС, оказывающих значительное влияние на ИИ, такими, как законы о конфиденциальности данных, законы об интеллектуальной собственности и антидискриминационные законы. Ожидается, что Закон будет вступать в силу поэтапно в течение двух лет, начиная с лета 2024 года.

Закон содержит определение системы ИИ как машинной системы, предназначенной для работы с различными уровнями автономии, которая может как проявлять адаптивность после развертывания и для достижения явных или неявных целей выводить из получаемых входных данных, так и генерировать выходные данные, такие как прогнозы, контент, рекомендации или решения, которые могут влиять на физическую или виртуальную среду. Это определение применяется независимо от того, каким образом система ИИ размещена на рынке или введена в эксплуатацию, в том числе в виде программного обеспечения с открытым исходным кодом. Система генеративного ИИ также может использоваться во множестве контекстов и интегрирована во множество других систем ИИ.

Закон распространяется на поставщиков (разработчиков), провайдеров, дистрибьюторов, импортеров ИИ, компании, использующие системы искусственного интеллекта («лица, развертывающие системы ИИ»). В нем устанавливаются подробные требования к ИИ общего назначения и «высокорискованному» ИИ, содержатся прямые запреты на определенные виды использования ИИ. В то же время Закон не распространяется на системы ИИ, предназначенные исключительно для научных исследований и разработок, а также на деятельность по НИОКР и тестированию систем/моделей ИИ вне реальных условий и до их выхода на рынок. Не входят в сферу регулирования Закона системы ИИ, используемые исключительно в военных целях, для обороны и национальной безопасности. Кроме того, свободные лицензии и лицензии с открытым исходным кодом не подпадают под действие Закона, если только они не относятся к запрещенным категориям или категориям высокого риска.

Предусмотренные Законом конкретные обязанности в отношении систем ИИ варьируются в зависимости от типа системы ИИ и, в частности, от целей, для которых она предназначена. Многие системы ИИ не подпадают ни под одну из категорий, к которым предъявляются особые требования в соответствии с Законом ЕС об ИИ – такие системы в значительной степени не регулируются Законом, но поставщикам и разработчикам необходимо учитывать влияние существующих законов, таких как GDPR. Кроме того, поставщики и разработчики всех систем ИИ подпадают под действие общего обязательство обеспечивать достаточный уровень грамотности в области ИИ среди сотрудников, которые взаимодействуют с ИИ.

Ключевыми категориями систем ИИ, установленными в Законе, являются:

Запрещенные системы ИИ, т. е. системы, которые, как считается, представляют неприемлемый риск для основных прав и свобод человека и находятся под запретом.

Искусственный интеллект с высокой степенью риска, т. е. системы, используемые для определенных целей, которые потенциально могут создать значительные риски (например, в контексте найма персонала) и подпадают под действие Закона.

Чат-боты и генеративный ИИ, в отношении которых Закон предусматривает относительно небольшое количество мер по обеспечению прозрачности и применения ИИ.

Искусственный интеллект общего назначения (генеративный ИИ), модели которого обладают значительной универсальностью, способны компетентно выполнять широкий спектр различных задач и могут быть интегрированы в различные последующие системы или приложения: к ним предъявляются различные специфические требования.

Если система ИИ подпадает под несколько вышеперечисленных категорий, будут применяться требования каждой категории. Например, поставщик чат-бота с ИИ, используемого для отбора кандидатов на работу, должен соответствовать следующим требованиям: обязательства, применимые как к системам высокого риска, так и к чат-ботам.

Запрещенный ИИ означает ограниченный набор практик, которые считаются особенно вредоносными и полностью запрещены, в том числе:

(a) распознавание эмоций на рабочем месте/в образовании – системы ИИ, предназначенные для определения эмоционального состояния людей в ситуациях, связанных с рабочим местом и образованием.

(b) нецелевое получение изображения для распознавания лиц из баз данных системы ИИ для распознавания данных с помощью нецелевого выделения изображения из Интернета или записи с камеры наблюдения.

(c) подсознательные техники и манипуляции, причиняющие значительный вред, т. е. системы ИИ, которые используют подсознательные техники или целенаправленные манипулятивные либо вводящие в заблуждение методы для искажения поведения человека или группы людей, в результате чего система ИИ причиняет (или может с большой вероятностью причинить) этому человеку, другому лицу или группе значительный вред.

(d) системы социальной оценки, которые оценивают отдельных лиц на основе их социального поведения или предполагаемых характеристик и приводят к пагубному или неблагоприятному обращению с людьми в другом контексте или являются неоправданными или непропорциональными;

(e) биометрические системы категоризации, основанные на биометрических данных отдельных лиц, позволяющие определить расовую принадлежность

[11] URL: https://artificialintelligenceact.eu/

человека, его политические взгляды, членство в профсоюзах, религиозные убеждения, сексуальную жизнь или сексуальную ориентацию;

(f) прогнозирующая полицейская деятельность, т. е. составление профилей лиц с целью прогнозирования вероятности совершения ими преступлений;

(g) идентификация в режиме реального времени для правоохранительных органов – использование систем удаленной идентификации в режиме реального времени в общественных местах для правоохранительных органов (за исключением определенных обстоятельств).

Законом предусмотрены штрафы за нарушение его положений. Максимальный штраф за несоблюдение Закона составляет более 35 млн евро, или 7% от общего годового оборота группы компаний по всему миру за предыдущий финансовый год, а по многим обязательствам максимальные штрафы превышают 15 млн евро, или 3% от общего годового оборота по всему миру.

Правовые аспекты применения ИИ: новые вызовы

В последние годы особенно остро встал вопрос о нормативно-правовом регулировании ИИ, процессов его создания и функционирования и, прежде всего, по вопросам ответственности ИИ, его разработчиков и пользователей.

Безопасность ИИ уже давно связывается с ответственностью за причинение вреда системами ИИ, т. е. вопросы определения юридической ответственности в случаях, когда решение, принятое с помощью алгоритма, оказывает негативное влияние на чью-либо жизнь, потенциальное преступное злоупотребление искусственным интеллектом и данными, а также использование ИИ в автономных системах вооружения. По мере развития технологий ИИ они будут бросать вызов основополагающим правовым принципам в соответствии с существующими концепциями частного права – в части ответственности при нарушении договоров или деликтной ответственности, а также ответственности при эксплуатации автономных машин и механизмов, робототехники, которую несет их владелец. Возможность привлечения к юридической ответственности в таких случаях часто заявляется главным препятствием для широкого внедрения ИИ[12].

Рассматриваются и аспекты регулирования ИИ с помощью этических норм, что свидетельствует о стремлении наделения систем ИИ правосубъектностью. С точки зрения правового регулирования безопасность ИИ складывается из нескольких элементов. Как отмечают одни исследователи [4], в отечественной правоприменительной практике все чаще возникают проблемы как договорной, так и деликтной ответственности (в т. ч. возмещения убытков) в связи с использованием роботов, вопросы конфиденциальности и безопасности в процессе такого использования, а также правового положения роботов, ответственности их разработчиков и пользователей, третьих лиц.

В сфере противоправной деятельности ИИ стал широко использоваться в преступных целях, а преступники все чаще используют искусственный интеллект и IT-технологии для совершения преступлений, где ИИ в виде роботов используется как орудие преступления – например, террористический акт, акт вандализма, хищение ядерных материалов и др. [5].

Следует отметить, что современная преступность в сфере компьютерных технологий с использованием ИИ становится многообразнее и изощреннее: в дополнение к существовавшим преступлениям в сфере компьютерной безопасности [6; 7; 8; 10], киберсквоттингу (регистрация доменных имен, похожих на уже зарегистрированные ранее товарные знаки или иные обозначения с целью последующей перепродажи или понуждению к покупке такого доменного имени правообладателем товарного знака) появились новые виды преступной деятельности в киберпространстве. В последние годы киберпреступность изменилась, перешла на «профессиональный» уровень – от одиночек к преступным сообществам.

Для организации кибератак, как отмечают эксперты[13], уже не нужно разбираться в зловредном программном обеспечении (ПО) и устройстве систем безопасности: рынок даркнета предлагает услуги на любые запросы – от простой DDoS-атаки до комплексных кампаний с применением уязвимостей нулевого дня. Сформировалась преступная практика – предоставление профессиональных услуг кибератак на аутсорсинге (“cybercrime as a service”).

Появился и новый вид преступлений – кражи моделей ИИ, что предполагает ситуацию, когда неавторизованные физические или юридические лица незаконно получают и используют модели ИИ, права на которые принадлежат другим лицам, без согласия их создателей. Наиболее популярные типы краж – это дистилляция модели и дообучение исходной модели на новом наборе данных с предварительным утаиванием способов получения исходной модели. Подобная ситуация произошла с известным французским стартапом Mistral – одной из самых производительных больших языковых моделей с открытым исходным кодом. Как правило, модели ИИ состоят из множества компонентов, что затрудняет отслеживание происхождения конкретных алгоритмов или фрагментов кода. Украденные модели подвергаются модификации: изменяя параметры, переобучая модели или добавляя в них новые

[12] AI in the UK: ready, willing and able? Report 2018, UK. URL: https://www.gov.uk/government/publications/ai-in-the-uk-ready-willing-and-able-government-response-to-the-select-committee-report

[13] Cybercrime as a service: как работает рынок киберпреступлений по заказу. URL: https://securitymedia.org/info/cybercrime-as-a-service-kak-rabotaet-rynok-kiberprestupleniy-po-zakazu.html

слои, злоумышленники усложняют установление прямой связи между украденной моделью и ее первоисточником[14].

Быстрое развитие ИИ (и генеративного ИИ в частности) создало целый лабиринт новых проблем с авторским правом – охраноспособность творческих результатов, созданных ИИ или с его непосредственным участием. ИИ общего назначения (AGI), который часто называют генеративным ИИ, занял видное место в связи с выпуском больших языковых моделей, таких как ChatGPT от OpenAI и Bard от Google. Но за этим последовали и судебные разбирательства и обвинения в нарушении законодательства об авторском праве. Например, компания Getty Images недавно объявила, что подает в суд на Stability AI, создателя AI Image Generator Stable Diffusion, за нарушение авторских прав.

В 2023 году суд в Вашингтоне (округ Колумбия) в деле *Талер против Бюро авторских прав США* постановил, что авторские права могут быть защищены только на произведения, созданные людьми, поскольку авторство человека является «основополагающим требованием авторского права», основанным на «веках устоявшегося понимания». Согласно решению суда, авторское право никогда не простиралось достаточно далеко, «чтобы защитить произведения, созданные с помощью новых форм технологии, функционирующих без какого-либо руководства со стороны человека, как настаивает здесь истец. Авторство человека является основополагающим требованием авторского права». Закон США об авторском праве 1909 года прямо предусматривал, что только человек может «защитить авторские права на свою работу». Аналогичным образом, апелляционный суд 9-го округа в 2018 году постановил, что «обезьяна, сделавшая селфи, не может подать в суд в соответствии с Законом об авторском праве за предполагаемое нарушение прав на фотографии самой себя, сделанные этой обезьяной, поскольку „все животные, поскольку они не являются людьми", не имели законного статуса в соответствии с этим законом». Талер не смог указать ни одного случая, «в котором суд признал бы авторское право на произведение, созданное не человеком».

В дополнение к указанным рискам и угрозам растет также обеспокоенность по поводу этических и социальных последствий применения систем ИИ в широком спектре социальных областей. Таким образом, бесспорно, что применение ИИ требует стандартизации и регулирования. К тому же под маркой «заслуживающего доверия ИИ» постулат, согласно которому ИИ должен соответствовать критериям прозрачности, законности, конфиденциальности, недискриминации и надежности. Под маркой «заслуживающего доверия ИИ» было сформулировано положение, согласно которому ИИ должен соответствовать критериям прозрачности, законности, конфиденциальности, недискриминации и надежности[15].

Вместе с тем развитие ИИ выявило ряд неожиданных проблем. Например, опыт лидера компании – разработчика ИИ OpenAI показал, что далеко не всегда частные компании, продвигающие новые революционное технологии, действуют в интересах не только своих акционеров, но и всего человечества. По оценке бывших членов совета директоров компании OpenAI [9], ИИ представляет собой новое оружие массового уничтожения. ИИ обладает огромным потенциалом для повышения производительности труда и благосостояния, но путь к этому лучшему будущему не лишен опасностей. Компания OpenAI была основана как смелый эксперимент по разработке всё более мощного ИИ, который предполагал приоритет общественного блага над получением прибыли. Однако, по мнению авторов, самоуправление не может надежно противостоять давлению стимулов, связанных с получением прибыли, и даже при всех преимуществах механизмов самоуправления, подобных тем, которые использует OpenAI, будет недостаточно. Поэтому крайне важно, чтобы государственный сектор активно участвовал в развитии технологии. Сейчас настало время правительственным органам всего мира заявить о себе. Только благодаря здоровому балансу рыночных сил и разумному регулированию мы можем надежно гарантировать, что эволюция ИИ действительно принесет пользу всему человечеству.

Угрозы, связанные с применением современных цифровых технологий, проявляются в наши дни особенно ярко. К примеру, как отмечает МИД РФ, «западные страны во главе с Вашингтоном продолжают наращивать именно наступательные ИКТ-потенциалы для проведения компьютерных операций против своих геополитических оппонентов, то есть России, Китая и других самостоятельных государств. О таких планах – под предлогом сдерживания противников – говорится в актуальных редакциях американских доктринальных документов, включая Стратегию кибербезопасности Пентагона и Международную стратегию в киберпространстве. О намерении вести борьбу с „недемократическими режимами" с помощью цифровых средств регулярно заявляют представители американской администрации» [2].

Заключение

ИИ способен помочь в решении самых сложных проблем, стоящих перед человечеством. Мир находится в разгаре технологической революции, которая коренным образом изменит человеческое общество. ИИ обещает преобразовать практически все аспекты эко-

[14] URL: https://securitymedia.org/news/rossiyskie-uchenye-zaregistrirovali-metod-vyyavleniya-krazhi-ii-modeley.html

[15] Jessica Newman, A Taxonomy of Trustworthiness for Artificial Intelligence. UC Berkeley Center for Long-Term Cybersecurity, 2023. URL:https://cltc.berkeley.edu/publication/a-taxonomy-of-trustworthiness-for-artificial-intelligence/

номики и общества, а его открывающиеся возможности носят трансформационный и глобальный характер. Однако огромные возможности технологий ИИ сопряжены с рисками, которые могут угрожать глобальной стабильности и подрывать общечеловеческие ценности, при этом создаваемые ИИ риски не признают национальных границ. Чтобы справиться с этими сложными и трудно прогнозируемыми рисками, смягчить потенциальные опасности, необходима разработка системных правовых мер и способов регулирования технологий и моделей ИИ, а также их применения как на национальном, так и на международном уровне.

Литература

1. Зиновьева Е., Исаева Т. Глобальный цифровой договор ООН: возможна ли прикладная реализация? // Международная жизнь. 2022. URL: https://interaffairs.ru/news/show/37601
2. Артур Люкманов: США давно отрабатывают методы гибридной войны против России. 25.05.2024, РИА Новости. URL: https://ria.ru/20240525/lyukmanov-1948182039.html
3. Klaus S., Jung C. Legal Aspects of "Artificial Intelligence" (AI). Information and Communication Technology Newsletter, 2019, No. 10. URL: https://www.swlegal.com/media/filer_public/ce/e4/cee498cc-910d-4af8-a020-b4063662b35/sw_newsletter_october_i_english.pdf
4. Харитонова Ю.С., Савина В.С., Паньини Ф. Гражданско-правовая ответственность при разработке и применении систем искусственного интеллекта и робототехники: основные подходы // Вестник Пермского университета. Юридические науки. 2022. № 4. С. 683—708.
5. Грачева Ю.В., Арямов А.А. Роботизация и искусственный интеллект: уголовно-правовые риски в сфере общественной безопасности // Актуальные проблемы российского права. 2020. № 6. С. 169—178.
6. Карцхия А.А., Макаренко Г.И. Правовые горизонты технологий искусственного интеллекта: национальный и международный аспект // Вопросы кибербезопасности. 2024. № 1. С. 6—17.
7. Карцхия А.А. Правовые аспекты современной киберпреступности // Правовая информатика. 2023. № 1. С. 83—92.
8. Карцхия А.А. Правовые аспекты современной кибербезопасности и противодействия киберпреступности // Вопросы кибербезопасности. 2023. № 1. С. 58—74.
9. Helen Toner, Tasha McCauley, AI firms mustn't govern themselves, say ex-members of Open AI's board. The Economist, 26th May 2024.
10. Карцхия А.А., Макаренко Г.И., Макаренко Д.Г. Правовые перспективы технологий искусственного интеллекта // Безопасные информационные технологии : сборник трудов XII Международной научно-технической конференции, 2023.

PUBLIC LAW (CONSTITUTIONAL LAW)

THE LEGAL FORMAT OF ARTIFICIAL INTELLIGENCE

Aleksandr Kartskhiia, *Dr.Sc. (Law), Professor at the Gubkin Russian State University of Oil and Gas, Moscow, Russian Federation.*
E-mail: *arhz50@mail.ru*

Keywords: *neural networks, technological sovereignty, breakthrough technologies, legal capacity, sustainable development, risk-oriented approach.*

Abstract

Purpose of the study: analysing the legal aspects of regulating Artificial Intelligence (AI) systems and technologies, new AI models for achieving the goals of sustainable development, identifying special features of using traditional legal categories in the AI field.

Methods used in the study: the comparative legal method, methods of analysis and synthesis in the process of historical genesis of AI, using the risk-oriented approach for assessing AI.

Study findings: the author comes to the conclusion that AI is capable of transforming virtually all aspects of economy and social relations in the society, and its newly opening opportunities are of transformational and global nature. However, outstanding possibilities offered by AI technologies are accompanied by risks that can threaten global stability and undermine universal human values. To overcome possible threats and risks and to decrease potential dangers, a strategy for developing systemic legal measures and ways for regulating AI technologies and models at the national and global scale is highly needed.

Research novelty: based on the risk-oriented approach, a legal format for the efficiency, stability and security of AI models is considered as well as determining AI's legal status, including from the standpoint of protecting mankind against uncontrolled impact of AI and invariability of guarantees for human rights and freedoms.

References

1. Zinov'eva E., Isaeva T. Global'nyi tsifrovoi dogovor OON: vozmozhna li prikladnaia realizatsiia? Mezhdunarodnaia zhizn', 2022. URL: https://interaffairs.ru/news/show/37601
2. Artur Liukmanov: SShA davno otrabatyvaiut metody gibridnoi voiny protiv Rossii. 25.05.2024, RIA Novosti. URL: https://ria.ru/20240525/lyukmanov-1948182039.html
3. Klaus S., Jung C. Legal Aspects of "Artificial Intelligence" (AI). Information and Communication Technology Newsletter, 2019, No. 10. URL: https://www.swlegal.com/media/filer_public/ce/e4/cee498cc-910d-4af8-a020-b4063662b35/sw_newsletter_october_i_english.pdf
4. Kharitonova Iu.S., Savina V.S., Pan'ini F. Grazhdansko-pravovaia otvetstvennost' pri razrabotke i primenenii sistem iskusstvennogo intellekta i robototekhniki: osnovnye podkhody. Vestnik Permskogo universiteta, Iuridicheskie nauki, 2022, No. 4, pp. 683–708.
5. Gracheva Iu.V., Ariamov A.A. Robotizatsiia i iskusstvennyi intellekt: ugolovno-pravovye riski v sfere obshchestvennoi bezopasnosti. Aktual'nye problemy rossiiskogo prava, 2020, No. 6, pp. 169–178.
6. Kartskhiia A.A., Makarenko G.I. Pravovye gorizonty tekhnologii iskusstvennogo intellekta: natsional'nyi i mezhdunarodnyi aspekt. Voprosy kiberbezopasnosti, 2024, No. 1, pp. 6–17.
7. Kartskhiia A.A. Pravovye aspekty sovremennoi kiberprestupnosti. Pravovaia informatika, 2023, No. 1, pp. 83–92.
8. Kartskhiia A.A. Pravovye aspekty sovremennoi kiberbezopasnosti i protivodeistviia kiberprestupnosti. Voprosy kiberbezopasnosti, 2023, No. 1, pp. 58–74.
9. Helen Toner, Tasha McCauley, AI firms mustn't govern themselves, say ex-members of Open AI's board. The Economist, 26th May 2024.
10. Kartskhiia A.A., Makarenko G.I., Makarenko D.G. Pravovye perspektivy tekhnologii iskusstvennogo intellekta. Bezopasnye informatsionnye tekhnologii : sbornik trudov XII Mezhdunarodnoi nauchno-tekhnicheskoi konferentsii, 2023.