ПРАВОВЫЕ ПРОБЛЕМЫ ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В РОССИИ

Карцхия А.А.[1], Макаренко Г.И.[2]

Ключевые слова: *нейронные сети, машинное обучение, безопасность искусственного интеллекта, правовая база.*

Аннотация

Цель работы: показать правовые проблемы при развитии и внедрении искусственного интеллекта в российской действительности.

Результат. Сделан обзор темы для России, США и Китая. Хотя Россия по использованию искусственного интеллекта находится на 10-м месте в мире, однако его внедрение идет быстрыми темпами. Авторам хотелось показать (и предостеречь), что внедрение того, что называют искусственным интеллектом, развивалось еще в СССР. Один из авторов еще в 1970 году создал лабораторию машинного проектирования для автоматического проектирования 13-слойных печатных плат бортовых вычислительных машин (авиакосмических комплексов). К 1980 году в СССР были сотни подразделений в самых разных областях техники, которые занимались автоматизацией проектирования и управления. Развитие автоматизации остановилось в России в связи с остановкой развития промышленности в стране — практически полностью было ликвидировано пассажирское самолетостроение, станкостроение, приборостроение и только в последние годы страна опомнилась и начала говорить о развитии промышленности. Правда, на примере самолетостроения мы видим, что даже давно испытанные и ранее выпускавшиеся пассажирские самолеты никак не начнут выпускаться.

На пути внедрения искусственного интеллекта — не только искусственные преграды в лице нерадивых чиновников, но и объективные обстоятельства: отсутствие правовой базы.

Практическая ценность: настоящая работа является дополнением статьи авторов «Правовые горизонты технологий искусственного интеллекта: национальный и международный аспект» (журнал «Вопросы кибербезопасности», № 1 за 2024 год) и может быть полезной при разработке правовой базы.

DOI: 10.21681/1994-1404-2024-1-4-19

Введение

Современный ландшафт искусственного интеллекта (ИИ), по мнению экспертов[3], показывает, что ИИ может быть отнесен к технологиям моделирования когнитивных функций человека, позволяющей компьютерам и машинам выполнять такие задачи, как решение проблем, принятие решений, понимание и воспроизведение естественного языка, распознавание паттернов и адаптация к изменяющейся среде. Взаимодействия человека и машины могут быть прямыми, когда люди взаимодействуют с интерфейсами ИИ, или косвенными, когда системы ИИ работают на втором плане для повышения производительности или принятия решений. ИИ может использовать преимущества других новых технологий и создавать синергию с ними. Например, блокчейн может записывать данные, которые когда-нибудь могут быть использованы ИИ для построения моделей и принятия обоснованных решений на основе проверенных данных. ИИ используется также для мониторинга транзакций как в децентрализованных финансах, так и в высокочастотной торговле традиционными финансовыми продуктами. Криптовалюты для оплаты могут применять вычислительную мощность ИИ. Инструменты и датчики Интернета вещей могут предоставлять огромные объемы данных, необходимых для обучения и обработки ИИ. Облачные технологии с их огромной вычислительной мощностью используются многими моделями и приложениями ИИ.

[3] Global Standards Mapping Initiative 4.0, November 2023. URL: https://gbbcouncil.org/wp-content/uploads/2023/11/GBBC-GSMI-4.0-Update-November-2023.pdf , pp. 11–12.

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

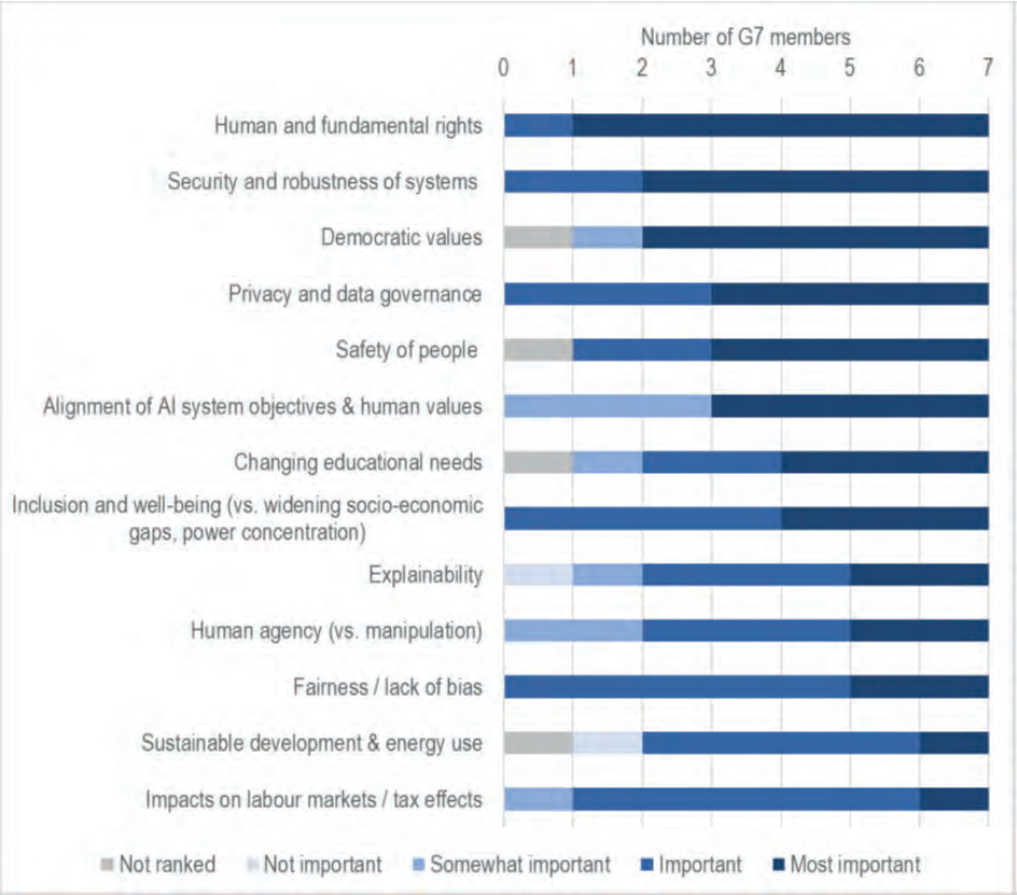
E-mail: arhz50@mail.ru

[2] ***Макаренко Григорий Иванович,*** *старший научный сотрудник ФБУ НЦПИ Минюста России, г. Москва, Российская Федерация.*

E-mail: t7920518@yandex.ru

Таблица 1.

G7 Hiroshima process on Artificial Intelligence (AI)



Вместе с тем выделяются и наиболее важные приоритеты ИИ, включая права человека и основные свободы, безопасность и надежность систем ИИ, демократические ценности, а также конфиденциальность и управление данными (см. табл. 1)[4].

В каждой из стран, входящих в лидеры разработок ИИ, в настоящее время активно реформируются направления и способы его правового регулирования[5], что подробнее мы рассмотрим далее.

Правовое регулирование сферы искусственного интеллекта в России

Для Российской Федерации развитие правового регулирования искусственного интеллекта связано, прежде всего, с ***Национальной стратегией развития искусственного интеллекта на период до 2030 года*** (далее — Стратегия), принятой в 2019 году и получившей существенные дополнения в 2024 году[6]. Стратегия провозглашает целями развития ИИ

[4] Необходимо особо ответить тем, кто боится искусственного интеллекта, кто верит, что будет восстание роботов и тому подобное. В ближайшие 50 лет такое невозможно представить. Любой существующий ИИ не способен создать ничего, чего бы не было, даже генеративный интеллект. Решение генерируется из анализа и комбинации существующих решений — вот почему любая программа генерирующего интеллекта обучается на основе большой базы решений. В этом у программ ИИ преимущество перед специалистом-человеком — человек выборку решений может выполнить за сутки и недели, в то время как программа ИИ делает это за секунды. Именно поэтому программы распознавания лиц нашли широкое распространение, так как задача сравнения лиц всегда работает с базой имеющихся лиц. Кстати, телевидение 15.03.2024 г. сообщило, что в России недавно было вынесено судебное решение об освобождение из заключения человека, которого программа ИИ определила как вероятного убийцу: сходство с портретом убийцы, произошедшее 20 лет назад, программа определила как 57%. Человек находился под следствием в заключении более года — виновата, разумеется, не программа, а следователи, которые при столь небольшом сходстве продержали человека в заключении более года. Хотелось бы знать, как наказаны следователи — если наказаны?

[5] Насколько преждевременны ожидания ИИ, показывает разоблачительная новость — компания Амазон закрывает магазины с технологией «взял и иди», где не нужно оплачивать товар на кассе, а просто уходить с ним. По задумке, компьютер сам распознавал, что Вы взяли, и автоматически списывал деньги с карты. Оказалось, что вместо ИИ за все отвечали свыше тысячи индусов, которые по камерам смотрели за покупателями и сами пробивали товар: так как технология ИИ слишком дорога для внедрения, в компании решили сэкономить с помощью дешевой рабочей силы. (Новость с Телеграмм-канала Левченко)

[6] Указ Президента РФ от 10.10.2019 № 490 (ред. от 15.02.2024) «О развитии искусственного интеллекта в Российской Федерации». URL: http://www.kremlin.ru/acts/bank/44731

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.

Понятие искусственного интеллекта получило уточнение в новой редакции Стратегии, где ***искусственный интеллект определяется*** 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. Комплекс технологических решений 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.

Сформулировано определение ***модели искусственного интеллекта,*** под которой понимается программа для электронных вычислительных машин (ее составная часть), предназначенная для выполнения интеллектуальных задач на уровне, сопоставимом с результатами интеллектуального труда человека или превосходящем их, использующая алгоритмы и наборы данных для выведения закономерностей, принятия решений или прогнозирования результатов.

Стратегия выделяет несколько новых понятий, включая:

- *большие генеративные модели* ИИ, которые способны интерпретировать (предоставлять информацию на основании запросов, например об объектах на изображении или о проанализированном тексте) и создавать мультимодальные данные (тексты, изображения, видеоматериалы и тому подобное) на уровне, сопоставимом с результатами интеллектуальной деятельности человека или превосходящем их;
- *большие фундаментальные модели* ИИ, т. е. модели, являющиеся основой для создания и доработки различных видов программного обеспечения, обученные распознаванию определенных видов закономерностей, содержащие не менее 1 млрд параметров и применяемые для выполнения большого количества различных задач;
- *перспективные методы* ИИ, т. е. методы, направленные на создание принципиально новой научно-технической продукции, в том числе в целях разработки универсального (сильного) искусственного интеллекта (автономное решение различных задач, автоматический дизайн физических объектов, автоматическое машинное обучение, алгоритмы решения задач на основе данных с частичной разметкой и (или) незначительных объемов данных, обработка информации на основе новых типов вычислительных систем, интерпретируемая обработка данных и другие методы);
- *доверенные технологии* ИИ — технологии, отвечающие стандартам безопасности, разработанные с учетом принципов объективности, недискриминации, этичности, исключающие при их использовании возможность причинения вреда человеку и нарушения его основополагающих прав и свобод, нанесения ущерба интересам общества и государства.

Стратегия отмечает, что ИИ является одной из самых важных технологий, которые доступны человеку в настоящее время: уже сейчас благодаря ИИ происходит рост мировой экономики, ускорение инноваций во всех областях науки, повышение качества жизни населения, доступности и качества медицинской помощи, качества образования, производительности труда и качества отдыха. Технологии ИИ являются областью международной конкуренции. Технологическое лидерство в области ИИ может позволить государствам достичь значимых результатов по основным направлениям социально-экономического развития. В конце 2010-х годов органы власти развитых стран стали уделять особое внимание развитию технологий ИИ. К настоящему времени более 60 стран разработали и утвердили собственные национальные стратегии развития ИИ.

Как указано в новой редакции Стратегии, в 2022—2023 годах в мире произошел новый скачок в развитии технологий ИИ благодаря совершенствованию больших генеративных моделей в области языка, изображений (включая видеоизображения) и звука. Большие фундаментальные модели уже сейчас способны писать программные коды по техническим заданиям, сочинять поэмы на заданную тему, давать точные и понятные ответы на тестовые вопросы различных уровней сложности, в том числе из образовательных программ. Модели ИИ за секунды создают изображения на любую тему по заданному текстовому описанию или наброску, что создает угрозу распространения запрещенной информации, нарушения авторских прав и генерации ошибочных сведений.

Искусственный интеллект окажет существенное влияние на экономический рост в мире. По оценкам экспертов, дальнейшее развитие больших генеративных моделей может вызвать резкое повышение производительности труда, которое приведет к увеличению мирового валового внутреннего продукта на 1—2% ежегодно и позволит повысить оплату труда специалистов во всех отраслях экономики за счет увеличения объема выпуска продукции (товаров, работ, услуг) и улучшения ее качества.

По итогам 2023 года, как отмечено в новой редакции Стратегии, в Российской Федерации созданы необходимые правовые условия для достижения целей, выполнения основных задач и реализации мер, предусмотренных настоящей Стратегией:

а) Правительство Российской Федерации утвердило Концепцию развития регулирования отношений в сфере технологий ИИ и робототехники до 2024 года[7];

б) сняты отдельные административно-правовые барьеры, препятствовавшие внедрению технологий ИИ в отдельных областях, включая здравоохранение, транспорт, государственно-частное партнерство и другие области;

в) принят Кодекс этики в сфере искусственного интеллекта[8], создана Комиссия по реализации Кодекса этики в сфере искусственного интеллекта и определены уполномоченные по этике в каждой организации, подписавшей данный Кодекс (по состоянию на ноябрь 2023 г. — 43 федеральных органа исполнительной власти, 17 органов исполнительной власти субъектов Российской Федерации, более 330 российских организаций и 23 иностранные организации присоединились к Кодексу этики в сфере искусственного интеллекта как стандарту, признанному на международном уровне);

г) сформирована система регулирования общественных отношений в области ИИ посредством публикации негосударственных актов рекомендательного характера («мягкое право»).

Российскими организациями создаются модели ИИ мирового уровня, в том числе в области генерации изображений, генерации и обработки текстов на русском и английском языках, медицины, генетики.

Как указывается в новой редакции Стратегии, изменение экономической ситуации, односторонние ограничительные меры недружественных иностранных государств и иные изменения рыночной конъюнктуры, которые произошли в 2022—2023 годах, определили новые вызовы для Российской Федерации:

а) нехватка вычислительных мощностей, недостаточное развитие отечественных решений в области ИИ, включая программно-аппаратные комплексы и электронную компонентную базу;

б) дефицит высококвалифицированных специалистов и инновационных разработок в области ИИ;

в) низкий уровень внедрения технологий ИИ в государственном управлении;

г) нехватка кадров для обеспечения массового внедрения технологий ИИ;

д) недостаточное субсидирование организаций, осуществляющих деятельность в области ИИ, и недостаток частных инвестиций в их развитие, в том числе на этапах предоставления венчурного финансирования, разработки концепции, проведения исследований, тестирования, промышленной разработки и эксплуатации технологий ИИ;

е) нормативные барьеры, препятствующие внедрению технологий ИИ в отдельных отраслях экономики, включая отсутствие методологической базы для обеспечения систем ИИ достоверными исходными данными;

ж) необходимость обеспечения безопасности при разработке и использовании технологий ИИ;

з) необходимость обеспечения защиты персональных данных и иной информации ограниченного доступа, объектов интеллектуальных прав при создании и обучении моделей ИИ;

и) ограничение доступа к технологиям ИИ в связи с недобросовестной конкуренцией со стороны недружественных иностранных государств и введением ими односторонних ограничительных мер;

к) возникновение в сфере разработки, создания и использования технологий ИИ новых типов угроз информационной безопасности, нехарактерных для других сфер применения информационных технологий;

л) дополнительные международные барьеры, препятствующие развитию ИИ в России и ограничивающие международное сотрудничество со стороны граждан и организаций недружественных иностранных государств.

Кроме того, новая редакция Стратегии определила цели и основные задачи развития ИИ, основные принципы развития и использования технологий ИИ, а также направления поддержки развития инфраструктуры и разработчиков технологий ИИ, стимулирование их внедрения и другие принципиальные вопросы регулирования ИИ.

Вместе с тем для обеспечения и защиты национальных интересов Российской Федерации от внешних и внутренних угроз, в том числе от недружественных действий иностранных государств, как указано в Стратегии национальной безопасности Российской Федерации[9], необходимо повысить эффективность использования имеющихся достижений и конкурентных преимуществ Российской Федерации с учетом долгосрочных тенденций мирового развития. 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, а также в целях научно-технологического развития России установлено развитие перспективных высоких технологий, таких как нанотехнологии, робототехника, медицинские, биологические, генной инженерии, информационно-коммуникационные, квантовые, ИИ, обработки больших данных, энергетические, лазерные, аддитивные, создания новых материалов, когнитивные, природоподобные технологии.

В целях повышения эффективности государственной научно-технической политики и обеспечения технологической независимости и конкурентоспособ-

[7] Распоряжение Правительства РФ от 19.08.2020 № 2129-р «Об утверждении Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники до 2024 года» // Собрание законодательства РФ. 2020. № 35. Ст. 5593.

[8] Кодекс этики в сфере искусственного интеллекта. URL: https://ethics.a-ai.ru/

[9] Указ Президента РФ от 02.07.2021 № 400 (ред. от 15.02.2024) «О Стратегии национальной безопасности Российской Федерации» // Собрание законодательства РФ. 2021. № 27 (часть II). Ст. 5351.

ности Российской Федерации, достижения ее национальных целей развития и реализации стратегических национальных приоритетов принят ***Указ Президента Российской Федерации от 02.11.2023 № 818 «О развитии природоподобных технологий в Российской Федерации»***[10], предусматривающий определение основных принципов и критериев отнесения технологий к природоподобным, а также разработку плана мероприятий, направленных на развитие природоподобных технологий в Российской Федерации и ее субъектах, в т. ч. на создание передовой научной инфраструктуры, формирование кадровых ресурсов и проведение научных исследований в этой сфере.

Природоподобные технологии представляют собой технологии, воспроизводящие системы и процессы живой природы в виде технических систем и технологических процессов, интегрированных в естественный природный ресурсооборот — так называемые конвергентные НБИКС-технологии (нано-, био-, информационные, когнитивные и социогуманитарные науки и технологии), которые, по мнению специалистов Курчатовского института[11], открывают возможность воспроизведения абсолютно всех систем и процессов живой природы и позволят создать гармоничную ноосферу, в которой три ее составляющие — биосфера, техносфера и общество — будут не конфликтовать, а дополнять друг друга, то есть будут конвергентны. Природоподобные технологии основаны на изучении образцов, объектов и процессов живой природы, осмыслении их механизмов и затем воспроизводстве в виде технических решений. В идеале возможно создание «биоискусственной клеточной системы», включая человека или его органов.

Эти технологии могут реализовываться в сочетании с технологиями ИИ. К примеру, японские специалисты смогли впервые в мире разработать технологию, которая с помощью генеративного искусственного интеллекта позволяет создавать изображения на основе обработки сигналов человеческого мозга (декабрь 2023 г.). Ранее проведенные исследования показали, что изображения, увиденные человеком, могут быть реконструированы на основе сигналов мозга, обработанных с помощью функциональной МРТ, но воспроизведению таким способом поддавались только изображения ограниченного круга. Теперь же разработанная программа преобразовывает сигналы мозга в числовые значения, на основе которых уже обученный искусственный интеллект воссоздает изображения[12].

В то же время российская ***Концепция развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники*** (далее — Концепция)[13], разработанная в целях определения основных подходов к трансформации системы нормативного регулирования в Российской Федерации для обеспечения возможности создания и применения таких технологий в различных сферах экономики с соблюдением прав граждан и обеспечением безопасности личности, общества и государства, исходит из того, что для развития технологий ИИ и робототехники необходимо создание регуляторной среды, комфортной для безопасного развития и внедрения указанных технологий, основанной на балансе интересов человека, общества, государства, компаний — разработчиков систем ИИ и робототехники, а также потребителей их товаров, работ, услуг.

К технологиям, основанным на использовании искусственного интеллекта, Концепцией (п. 5) отнесены:

а) компьютерное зрение;

б) обработка естественного языка;

в) распознавание и синтез речи;

г) интеллектуальная поддержка принятия решений;

д) перспективные методы искусственного интеллекта.

Перспективными методами искусственного интеллекта признаются:

а) автономное решение различных задач;

б) автоматический дизайн физических объектов;

в) автоматическое машинное обучение;

г) алгоритмы решения задач на основе данных с частичной разметкой и (или) незначительных объемов данных;

д) обработка информации на основе новых типов вычислительных систем;

е) интерпретируемая обработка данных;

ж) другие методы.

В Концепции отмечается, что повышение степени автономности систем ИИ и робототехники, снижение контроля человека за процессом их применения, не полностью прозрачный процесс принятия решений создают общественный запрос на регуляторные ограничения применения систем ИИ и робототехники. В настоящее время в мире отсутствуют единые подходы к регулированию технологий ИИ и робототехники, что связано с наличием ряда проблем, не имеющих однозначного решения.

В Концепции формулируется ***российская правовая модель регулирования искусственного интеллекта*** в соответствии с Национальной стратегией развития искусственного интеллекта на период до 2030 г., которая предусматривает следующие основные направления создания комплексной системы регулирования общественных отношений, возникающих в связи с развитием и внедрением технологий ИИ:

[10] Собрание законодательства РФ. 2023. № 45. Ст. 8035.

[11] URL: https://nauka.tass.ru/nauka/19185825?ysclid=lq7qhbicfb382166577

[12] URL: https://nauka.tass.ru/nauka/19556253

[13] Утв. Распоряжением Правительства РФ от 19.08.2020 № 2129-р «Об утверждении Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники до 2024 года» // Собрание законодательства РФ. 2020. № 35. Ст. 5593.

а) обеспечение благоприятных правовых условий (в том числе посредством создания экспериментального правового режима) для доступа к данным, преимущественно обезличенным, включая данные, собираемые государственными органами и медицинскими организациями;

б) обеспечение особых условий (режимов) для доступа к данным, включая персональные, в целях проведения научных исследований, создания технологий ИИ и разработки технологических решений на их основе;

в) создание правовых условий и установление процедур упрощенного тестирования и внедрения технологических решений, разработанных на основе ИИ, а также делегирования информационным системам, функционирующим на основе ИИ, возможности принятия отдельных решений (за исключением решений, которые могут ущемлять права и законные интересы граждан), в том числе при исполнении государственными органами государственных функций (за исключением функций, направленных на обеспечение безопасности населения и государства);

г) устранение административных барьеров при экспорте продукции (работ, услуг) гражданского назначения, созданной на основе ИИ;

д) создание единых систем стандартизации и оценки соответствия технологических решений, разработанных на основе ИИ, развитие международного сотрудничества Российской Федерации по вопросам стандартизации и обеспечение возможности сертификации продукции (работ, услуг), созданной на основе ИИ;

е) стимулирование привлечения инвестиций посредством совершенствования механизмов совместного участия инвесторов и государства в проектах, связанных с разработкой технологий ИИ, а также предоставления целевой финансовой поддержки организациям, осуществляющим деятельность по развитию и внедрению технологий ИИ (при условии, что внедрение таких технологий повлечет за собой существенные позитивные эффекты для отраслей экономики Российской Федерации);

ж) разработка этических правил взаимодействия человека с ИИ.

Такие направления должны стать основными ориентирами при создании комплексной системы регулирования общественных отношений, возникающих в связи с развитием и внедрением технологий ИИ и робототехники.

В Концепции предусматривается, что с учетом экономической и социальной значимости применения технологий ИИ и робототехники в различных сферах их разработка и эксплуатация не должны ограничиваться регуляторными мерами, за исключением случаев, связанных с высоким риском причинения вреда жизни и здоровью граждан. Не допускается также применение технологий ИИ и робототехники, представляющих явную угрозу обороне страны и безопасности государства.

Для выработки конкретных регуляторных решений требуется использовать риск-ориентированный подход, основанный на оценке размера потенциального вреда указанным ценностям с учетом вероятности его наступления по сравнению с потенциальным положительным эффектом от внедрения технологий ИИ и робототехники, необходимости принятия мер по минимизации соответствующих рисков.

Сам факт использования систем ИИ и робототехники не должен являться основанием для установления регуляторных ограничений.

Следует поддерживать развитие регулирования, вырабатываемого и приводимого в исполнение силами участников рынка (саморегулирование), включая принятие и использование документов национальной системы стандартизации, кодексов (сводов) этических правил и иных документов саморегулируемых организаций, а также иных инструментов.

Учитывая принципиальную сложность регулируемой сферы правоотношений, для выработки режима регулирования технологий ИИ и робототехники требуется активное вовлечение представителей компаний — разработчиков систем ИИ и робототехники, научно-исследовательских организаций в процесс экспертной проработки соответствующих нормативных правовых актов. В дальнейшем может также потребоваться уточнение отдельных норм законодательства в целях нормативного правового регулирования новых видов правоотношений.

В российском законодательстве также предусмотрены специальные экспериментальные правовые режимы регулирования ИИ. Так, Федеральным законом от 24.04.2020 № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации — городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона «О персональных данных»[14] предусмотрена возможность установления специального регулирования в целях создания необходимых условий для разработки и внедрения технологий ИИ в городе Москве, а также последующего возможного использования результатов применения ИИ.

Важным вопросом является финансовое стимулирования разработок ИИ, в т. ч. в целях предоставления субсидии из федерального бюджета на поддержку разработки и коммерциализации новых технологий пилотных проектов апробации технологий ИИ в приоритетных отраслях установлены критерии соответствия технологии ИИ[15]. В частности, финансирование предоставляется, если проект удовлетворяет критерию базовой технологии проекта, а его мероприятия предусматривают создание, и (или) развитие, и (или) внедрение

[14] Собрание законодательства РФ. 2020. № 17. Ст. 2701.

[15] Приказ Минэкономразвития России от 29.06.2021 № 392 «Об утверждении критериев определения принадлежности проектов к проектам в сфере искусственного интеллекта». URL: http://pravo.gov.ru, 29.07.2021.

не менее чем одной из технологий ИИ, а также если его мероприятия направлены на решение технологических задач, установленных перечнем технологических задач, на реализацию которых может быть направлен проект в сфере ИИ, приведенным в приложении к настоящим Критериям.

Для целей определения соответствия проекта критерию базовой технологии к перспективным методам ИИ относятся автономная работа физических машин (робототехника) и обработка информации на основе новых типов специализированных вычислительных систем для задач ИИ.

В то же время необходимо учитывать, что безопасность, надежность и устойчивость применения современных технологий, особенно ИИ, ставится во главу угла регулирования этой сферы, в т. ч. в сфере сбора и обработки информации.

Учитывая существенный объем информации, доступный для потребления человеку, размещаемый в социальных сетях и иных площадках в Интернете, для упрощения его получения используются «рекомендательные алгоритмы», которые предлагают пользователю контент на основе его интересов и предпочтений. Однако подобные технологии не всегда добросовестно используются владельцами социальных сетей и иных информационных ресурсов. Так, под видом рекомендации пользователю может быть умышленно представлена информация, вводящая его в заблуждение или нарушающая законы Российской Федерации (распространение «фейковых» новостей, скрытая реклама и т. п.)[16].

В связи с этим был принят и с 1 октября 2023 года вступил в силу Федеральный закон от 31.07.2023 № 408-ФЗ «О внесении изменения в Федеральный закон «Об информации, информационных технологиях и о защите информации»[17], который устанавливает новые требования для владельцев сайтов в сети Интернет и мобильных приложений (за исключением операторов государственных информационных систем, государственные органы и органы местного самоуправления), применяющие технологии предоставления информации на основе сбора, систематизации и анализа сведений, относящихся к предпочтениям пользователей сети Интернет («рекомендательные технологии»). Владелец такого сайта или приложения обязан не допускать применения рекомендательных технологий, которые нарушают права и законные интересы граждан и организаций, а также применения их в целях предоставления информации с нарушением законодательства Российской Федерации. Обязательно также размещение на сайте или в приложении правил применения рекомендательных технологий, которые должны включать в себя описание процессов и методов сбора, систематизации, анализа сведений, относящихся к предпочтениям пользователей сети Интернет, способов осуществления таких процессов и методов, перечень сведений, относящихся к предпочтениям пользователей сети Интернет, которые используются для предоставления информации с применением рекомендательных технологий, а также источники получения таких сведений. В случае установления факта неисполнения владельцем информационного ресурса, на котором применяются рекомендательные технологии, указанных обязанностей, ресурс может быть заблокирован Роскомнадзором.

Актуальное регулирование искусственного интеллекта в США

В марте 2023 г. Президент США утвердил новую редакцию Национальной стратегии в области кибербезопасности[18], которая определяет, что защита критически важной инфраструктуры США стала приоритетом национальной безопасности. Инициатива направлена на то, чтобы переложить часть бремени снижения рисков кибербезопасности с конечных пользователей и операторов критически важной инфраструктуры на предприятия частного сектора, которые лучше всего расположены для достижения значимых успехов в области безопасности и отказоустойчивости. В Стратегии также подчеркивается необходимость изменения стимулов в пользу долгосрочных инвестиций частного сектора. Стратегия построена на пяти основных задачах:

1) защита критически важной инфраструктуры;

2) выявление и уничтожение субъектов угроз;

3) формирование рыночных механизмов повышения безопасности и устойчивости;

4) инвестиции в устойчивое будущее;

5) налаживание международных партнерских отношений для достижения общих целей.

Каждый компонент содержит конкретные стратегические цели, разработанные на основе предыдущих программ, и направляющие усилия по реализации государственных структур и организаций частного сектора.

Стратегия сформулировала новую волну регулирования, которая направлена на внедрение новой парадигмы регулирования кибербезопасности в секторах критически важной инфраструктуры путем перехода от добровольных руководящих принципов к обязательным кибернетическим правилам, которые, как признается в Стратегии, потребуют определенных законодательных действий. Движущей силой этой инициативы является требование более целенаправленного, более скоординированного и более обеспеченного ресурсами подхода к киберзащите.

В Стратегии также признаются повышенные риски в нынешнюю эпоху глобальной цифровизации и углубления цифровой зависимости, ускоряемого появлением новых технологий. Стремительный технологический

[16] Разъяснение Прокуратуры Московской области от 16.11.2023 «Упорядочено применение рекомендательных технологий на сайтах в Интернете». URL: https://epp.genproc.gov.ru/web/proc_50, 16.11.2023.

[17] Собрание законодательства РФ. 2023. № 32 (Часть I). Ст. 6140.

[18] URL: https://www.whitehouse.gov/wp-content/uploads/2023/03/National-Cybersecurity-Strategy-2023.pdf

прогресс также вынуждает секторы критически важной инфраструктуры бороться с рисками конвергенции информационных технологий и операционных технологических систем, которые должны проектироваться и защищаться совершенно разными способами. Сложная геополитическая обстановка усугубляет эти риски, поскольку число киберугроз для критически важной инфраструктуры, спонсируемых государством, растет. Хотя конкретные способы реализации Стратегии не определены, оперативная реализация целей будет иметь ключевое значение в развивающемся мире, где угрозы могут опережать регулирование и законотворчество. Администрация Байдена сослалась на некоторые всеобъемлющие принципы в дополнение к обязательным нормативным актам, например, продуманная безопасность как основной бизнес-принцип, оперативная доступность, позволяющая избежать системных сбоев, содействие гармонии нормотворчества в разных юрисдикциях.

Введен новый вид страхования — киберстрахование как относительно новый вид страхования, который покрывает различные виды ответственности или прямые убытки от событий, связанных с электронной деятельностью и системами, а также предполагающий партнерство между правительством и страховой отраслью для поддержки системам кибербезопасности коммерческих организаций в соответствии с национальными целями.

При всех ее преимуществах и перспективных инициативах анализ и реализация Стратегии сопряжены с некоторыми трудностями, включая формулирование и реализацию требований отчетности для частных компаний, столкнувшимся с киберинцидентами. Федеральное правительство установило меры для улучшения национальной кибербезопасности и возможностей перед лицом усиливающихся угроз кибербезопасности. Среди них — рекомендации Агентства кибербезопасности и инфраструктурной безопасности (входит в состав Министерства внутренней безопасности США) по спецификации программного обеспечения и обновлению межотраслевых показателей кибербезопасности, предлагаемые требования Комиссии по ценным бумагам и биржам США в отношении рисков кибербезопасности, меморандум Агентства по охране окружающей среды США в отношении систем общественного водоснабжения и распространение директив Управления транспортной безопасности, ориентированных на безопасность трубопроводов, на авиационный и железнодорожный секторы.

Кроме того, владельцам и операторам критически важной инфраструктуры предписано предпринять ряд ключевых мер, включая:

- участие в разработке официальных требований и стандартов образование регулирующих органов имеет решающее значение;
- координация деятельности внутренних IT и кадровых структур безопасности, комплаенса и юридическими подразделениями;
- использование рекомендаций, стандартов, лучших практик и непрерывное обучение для укрепления кибербезопасности.

Вместе с тем обращает на себя внимание откровенно агрессивный характер документа, который, в частности, легитимизирует наступательные кибероперации в качестве превентивной или ответной меры для подавления хакерских группировок и иных киберсил в информационном пространстве третьих стран. Этот подход в целом укладывается в логику известных американских концепций «постоянного воздействия» и «наступательной обороны», продвигаемых Агентством национальной безопасности и киберкомандованием Вооруженных сил США. Наглядным примером является ставка на проведение экстерриториальных расследований киберпреступлений. На это нацелена не ратифицированная Россией так называемая Будапештская конвенция. Она призвана легитимизировать право США без уведомления властей других стран вторгаться в их информационное пространство для сбора «цифровых улик»[19]. Характерно, что международные аналитические исследования[20] подтверждают доминирование наступательных методов в стратегиях кибербезопасности таких государств, как США, Великобритания и Украина.

30 октября 2023 г. президент Байден издал новый Указ о безопасном и заслуживающим доверия искусственном интеллекте *(Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence)*[21], 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 представляющих серьезную угрозу национальной безопасности, национальной экономической безопасности или национальному 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.

---

[19] Интервью заместителя секретаря Совета безопасности РФ О. Храмова «Российской газете» 11 октября 2023 года. URL: http://www.scrf.gov.ru/news/allnews/3573/

[20] Стратегии кибербезопасности. Аналитический отчет, InfoWatch, 2022. URL: https://www.infowatch.ru/analytics/analitika/strategii-kiberbezopasnosti

[21] 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/

Этот указ включает более сотни конкретных директивных указаний, 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 2024 г.

Правовое регулирование искусственного интеллекта в КНР

За последние несколько лет Китай добился значительных успехов в своих усилиях стать технологической сверхдержавой, постоянно прилагая усилия для утверждения себя в качестве ведущего мирового производителя ИС.

Страна превратилась из экономики с низкой заработной платой в высокотехнологичную державу. Фактически, по данным Всемирной организации интеллектуальной собственности (ВОИС), на Китай пришлось 46,8% всех патентных заявок по всему миру в 2023 году[22] , что свидетельствует о его стремлении избавиться от своего прошлого имиджа.

13 июля 2023 года правительство Китая опубликовало правила по генеративному искусственному интеллекту, Временные меры по управлению службами генеративного искусственного интеллекта (далее — Временные меры ГАИ)[23], которые вступили в силу 15 августа 2023 года. Целью Временных мер ГАИ является регулирование генеративного ИИ, который в первую очередь предназначен для создания контента, и они являются последним дополнением к формирующейся системе регулирования искусственного интеллекта в КНР, которая уже включает ряд специфичных для ИИ и местных законов.

Китайское правительство с самого начала начало поддерживать свою индустрию ИИ на национальном уровне. В 13-м пятилетнем плане Пекина (2016—2020) ИИ определен как ключевой для достижения целей экономического роста. В 2017 года китайское правительство представило свое вйдение развития ИИ в Китае, опубликовав свой план развития ИИ следующего поколения. В Плане представлена комплексная стратегия Пекина по сосредоточению ИИ в усилиях Китая по социально-экономическому развитию — индустрия ИИ, которая сделает Китай мировым лидером в области ИИ к 2030 году.

Со временем в рамках этих более широких стратегий в Китае на различных уровнях были собраны воедино различные законы об ИИ. На региональном уровне первый провинциальный закон Китая о разработке ИИ вступил в силу 1 октября 2020 года с принятием Шанхайских правил содействия развитию ИИ индустрии, которые направлены на продвижение индустрии ИИ на муниципальном уровне в Шанхае. Вскоре после этого правительство Шэньчжэня приняло аналогичный закон, Положение о продвижении индустрии искусственного интеллекта в Особой экономической зоне Шэньчжэня, который вступил в силу 1 ноября 2022 года.

Пекин также начал закладывать основу для конкретного решения проблемы генеративных систем ИИ. Положения об управлении алгоритмическими рекомендациями информационной службы Интернета («Положения об алгоритмических рекомендациях»), которые устанавливают структуру управления для регулирования систем рекомендаций, вступили в силу 1 марта 2022 года. Кроме того, правительство КНР выпустило целевые правила для генеративного ИИ через Положения об управлении глубоким синтезом интернет-сервиса («Положения о глубоком синтезе»), которые вступили в силу 10 января 2023 года и применяются к результатам глубокой подделки с помощью технологии ИИ. Временные меры ГАИ основаны на Положениях о глубоком синтезе, предоставляя более конкретные правила и принудительные меры конкретно для генеративного ИИ.

Вместе с Положениями об алгоритмических рекомендациях и Положениями о глубоком синтезе Временные меры ГАИ составляют основную правовую основу для соблюдения нормативных требований и надзора за ИИ индустрией в Китае.

Временные меры ГАИ отличаются от других законов тем, что конкретно регулируют использование генеративной технологии ИИ, определяемой как «модели и связанные с ними технологии, которые обладают способностью генерировать контент, такой как текст, изображения, аудио и видео», для предоставления услуг по генерации контента населению на материковой части КНР. По сравнению с положениями Deep Synthesis, генеративные технологии ИИ, подпадающие под временные меры GAI, охватывают нечто бóльшее, чем генеративные технологии, основанные на алгоритмах, и могут синтезировать технологии, включая также модели и системы, основанные на правилах.

Временные меры ГАИ применяются к поставщикам генеративных услуг ИИ, определяемым как организации и частные лица, которые используют генеративные технологии ИИ для предоставления генеративных услуг ИИ, включая предоставление этих услуг через интерфейсы прикладного программирования (API). Указано также, что «пользователи» генеративных услуг ИИ определяются как организации и от-

[22] IP Facts and Figures, WIPO, 2023. URL: https://www.wipo.int/en/ipfactsandfigures/patents

[23] URL: http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm

дельные лица, которые используют генеративные услуги ИИ для создания контента. Временные меры ГАИ охватывают предоставление генеративных услуг ИИ населению косвенно посредством деловых соглашений. Однако учреждения, которые разрабатывают и применяют генеративную технологию ИИ, но не предоставляют генеративные услуги ИИ населению, освобождаются от ответственности. Кроме того, Временные меры ГАИ устанавливают экстерриториальную сферу действия, уточняя, что они применяются к предоставлению услуг населению на материковой территории КНР, потенциально распространяя их применение на частных лиц и организации за пределами Китая, которые предоставляют генеративные услуги ИИ лицам в КНР. Этот нюанс дополняется другим положением, в котором говорится, что несоблюдение Временных мер ГАИ и других законов поставщиками генеративного ИИ за пределами КНР приведет к уведомлению соответствующих учреждений о принятии технических мер и других необходимых мер для борьбы с ними. Примечательно, что Временные меры ГАИ имеют значительно меньшую сферу применения по сравнению с ранним проектом, который также применялся бы к исследованиям, разработке и использованию продуктов с генеративными функциями ИИ. Однако эта формулировка не упоминается во Временных мерах ГАИ.

Временные меры ГАИ устанавливают ряд общих требований к предоставлению и использованию генеративных услуг ИИ:

- Уважение общественной морали и нравственности Китая и отстаивание «основных социалистических ценностей».
- «Эффективные меры» должны применяться во время разработки алгоритмов, отбора обучающих данных, генерации и оптимизации моделей, предоставления услуг и других процессов для предотвращения дискриминации по таким факторам, как раса, этническая принадлежность, религиозные убеждения, национальность, регион, пол, возраст, профессия или состояние здоровья.
- Уважение прав интеллектуальной собственности, коммерческой этики и защита коммерческой тайны, а также запрет на использование в целях монополии и недобросовестной конкуренции.
- Уважение законных прав и интересов других лиц и запрет на создание угрозы физическому и психологическому благополучию других лиц или нарушение их прав и интересов, включая их имидж, репутацию, честь, неприкосновенность частной жизни и личную информацию.
- Необходимо использовать эффективные меры для повышения прозрачности, точности и надежности услуг генеративного ИИ.

Кроме того, предоставление и использование генеративных услуг ИИ не должно генерировать контент, который может привести к следующим негативным результатам:

- подстрекательство к подрыву национального суверенитета или свержению социалистической системы;
- создание угрозы национальной безопасности и интересам или нанесение ущерба имиджу китайской нации;
- разжигание сепаратизма или подрыв национального единства и социальной стабильности;
- пропаганда терроризма или экстремизма;
- пропаганда этнической ненависти и дискриминации, насилия и непристойностей, а также фальшивой и вредной информации.

Временные меры ГАИ предъявляют более конкретные эксплуатационные требования к поставщикам услуг генеративного ИИ, чем первая версия. Эти операционные требования касаются целого ряда вопросов, таких как разработка модели, управление данными, стандарты обслуживания и модерация контента, и включают следующее.

- Данные для обучения. Поставщики должны использовать данные и базовые модели из «законных источников» и применять «эффективные меры» для повышения качества, достоверности, точности, объективности и разнообразия данных для обучения.
- Права на неприкосновенность частной жизни. Провайдеры несут ответственность как обработчики личной информации и должны получать согласие при использовании личной информации, если не применяется исключение, и соблюдать правила конфиденциальности, включая сбор и минимизацию данных, хранение данных и индивидуальные права в отношении доступа, исправления и удаления.
- Маркировка данных. Поставщики услуг должны установить четкие, конкретные и практичные правила маркировки, которые соответствуют требованиям Временных мер ГАИ в процессе исследований и разработок для генеративного ИИ.
- Модерация контента и маркировка. Провайдеры несут ответственность как производители информационного онлайн-контента и должны маркировать созданный контент в соответствии с требованиями Положений Deep Synthesis, оперативно устранять «незаконный контент» с помощью «эффективных мер», таких как прекращение генерации или передачи контента, его удаление и исправление с помощью обучения оптимизации модели, а также сообщать о проблеме в соответствующие органы.
- Взаимодействие с пользователями и жалобы. Провайдеры должны заключать соглашения об обслуживании с пользователями для уточнения прав и обязанностей, использовать «эффективные меры» для предотвращения зависимости или чрезмерного увлечения несовершеннолетними пользователями и разработать механизмы рассмотрения жалоб пользователей.

– Оценка безопасности. Поставщики генеративных услуг ИИ, обладающих «свойствами общественного мнения» или «способностью к социальной мобилизации», должны проводить оценки безопасности и соблюдать требования к алгоритмической подаче документов в соответствии с Положениями алгоритмических рекомендаций.

За нарушения соответствующие регулирующие органы должны налагать штрафы в соответствии с применимыми законами и административными постановлениями, включая законы КНР о кибербезопасности, о безопасности данных, о защите личной информации и о научно-техническом прогрессе, если нарушение не касается определенной области, соответствующие регулирующие органы могут выносить предупреждения, публиковать критические замечания и предписывать поставщикам соблюдать требования законодательства. При «серьезных обстоятельствах» или в случае отказа в таких распоряжениях власти могут распорядиться о приостановке работы услуг генеративного ИИ. Однако, в отличие от черновой версии, которая предусматривала бы денежные штрафы в размере от 10 000 до 100 000 юаней за нарушения, Временные меры ГАИ не предусматривают денежных штрафов.

Поставщики услуг также должны быть готовы сотрудничать с инспекциями регулирующих органов, имея возможность объяснить обучающие данные, включая их источники, модели, типы, правила маркировки и алгоритмы, а также предоставить любую другую необходимую помощь. Органы власти, участвующие в регулировании генеративного ИИ, также обязаны сохранять конфиденциальность конфиденциальных данных, таких как коммерческая тайна и личная информация, полученная в ходе выполнения своих обязанностей, и не должны разглашать или незаконно передавать эти данные третьим лицам.

Современное законодательство ЕС в сфере ИИ

В марте 2024 г. Европарламент принял самый амбициозный закон, регулирующий ИИ, текст которого согласовали все страны ЕС. Закон об ИИ ЕС (*EU Artificial Intelligence Act*)[24] (далее — Закон) занимает более 200 страниц и применяется как к поставщикам, так и к пользователям технологий, основанных на ИИ, в частном и государственном секторах. Как и в других законодательных актах ЕС, связанных с данными, Закон также применяется экстерриториально к компаниям и организациям за пределами ЕС.

Закон определяет ***систему искусственного интеллекта*** (в англоязычной версии) как *машинную систему, предназначенную для работы с различными уровнями автономии, которая может как проявлять адаптивность после развертывания и которая для достижения явных или неявных целей выводит 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 соответствует определению Организации экономического сотрудничества и развития (ОЭСР).

Согласно определению, ключевой характеристикой, отличающей системы ИИ от традиционного программного обеспечения, является то, что система ИИ делает выводы на основе входных данных (*«выводит из получаемых входных данных и генерирует выходные данные»*). Это призвано подчеркнуть способность систем ИИ создавать модели и (или) алгоритмы из входных данных. Первоначально Закон предполагал исключить системы, основанные на правилах, которые определяются исключительно физическими лицами, для выполнения автоматических процессов, из сферы действия Закона об ИИ. По определению, возможности систем ИИ должны выходить за рамки базовых операций по обработке данных и пониматься скорее как обучение, рассуждения или моделирование. Определение в Законе также предполагает, что системы ИИ *«предназначены для работы с различными уровнями автономии»*. Соответственно, должна быть определенная степень независимости действий системы от людей. Другими словами, система должна быть способна работать без вмешательства человека. Характеристика *«адаптивности»* предназначена для выражения способности системы ИИ (продолжать) изучать саму себя и, таким образом, постоянно меняться.

Технологии ИИ, которые представляют явный риск для основных прав человека (биометрические системы категоризации на основе чувствительных характеристик, социальный рейтинг или ИИ, используемый для манипулирования поведением человека), будут запрещены в Европе. Системы ИИ, считающиеся высокорискованными, используемые, например, в критической инфраструктуре, образовании, здравоохранении, правоохранительных органах, управлении границами или на выборах, должны будут соответствовать строгим требованиям.

Закон определяет *модели искусственного интеллекта общего назначения (GPAI)*. Среди прочих требований, разработчики моделей GPAI должны создавать и обновлять техническую документацию модели, включающую определенные минимальные элементы, а также предоставлять подробную информацию и документацию поставщикам, которые интегрируют эти модели в свои системы ИИ, соблюдать законы об авторском праве ЕС и публиковать «достаточно» подробное изложение контента, используемого для обучения модели GPAI.

Отдельно выделяются *модели искусственного интеллекта общего назначения с системным риском*. Закон налагает повышенные требования на поставщиков моделей GPAI «с системными рисками», которые включают требования к проведению оценки модели, включая состязательное тестирование модели, для оценки и смягчения возможных системных рисков на уровне

[24] URL: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

ЕС, а также обеспечения адекватной защиты в области кибербезопасности.

Предусмотрены *исключения из систем искусственного интеллекта высокого риска.* Наиболее жесткие требования Закона распространяются на системы ИИ «высокого риска». В Законе определены два типа систем ИИ, представляющих высокий риск:

(1) системы ИИ, предназначенные для использования в качестве продуктов (или компонентов безопасности продуктов), на которые распространяется специальное законодательство ЕС, перечисленное в приложении II к Закону, и

(2) системы ИИ, используемые для целей, перечисленных в Приложении III к Закону, таких как определенное использование систем удаленной биометрической идентификации и определенных систем ИИ, используемых для правоохранительных органов.

Есть и исключение из этого требования: если система искусственного интеллекта, подпадающая под действие Приложения III, «не представляет значительного риска нанесения вреда здоровью, безопасности или основным правам физических лиц», поставщик может задокументировать это и на этом основании исключить систему из обязательств Закона в отношении таких систем. Органы по надзору за рынком уполномочены оценивать системы, которые, по их мнению, были неправильно классифицированы, и предписывать меры по исправлению положения. Провайдеры также будут подвергнуты штрафам, если орган по надзору за рынком установит, что провайдер неправильно классифицировал свою систему ИИ, чтобы обойти применение обязательств к системам ИИ высокого риска.

Организации, внедряющие системы, которые регулируются публичным правом, частные операторы, предоставляющие государственные услуги, и (за некоторыми исключениями) операторы, внедряющие системы ИИ высокого риска для оценки кредитоспособности физического лица, установления кредитного рейтинга физического лица или оценки рисков и цен в связи со страхованием жизни или здоровья физического лица, должны выполнить оценку воздействия на основные права перед внедрением системы ИИ высокого риска. Эти требования включают:

- процессы разработчика, в которых система ИИ высокого риска будет использоваться в соответствии с ее назначением;
- описание периода времени и частоты, с которой предполагается использовать систему ИИ высокого риска;
- категории физических лиц и групп, которые могут пострадать от его использования в конкретном контексте;
- конкретные риски причинения вреда, которые могут повлиять на категории лиц или группу лиц, определенных как подверженные влиянию, принимая во внимание информацию, предоставленную поставщиком услуг в соответствии с его обязательствами по прозрачности в соответствии со статьей 13;
- описание реализации мер по надзору за персоналом в соответствии с инструкциями по использованию; и
- меры, которые необходимо предпринять в случае реализации этих рисков, включая их механизмы внутреннего управления и подачи жалоб.

Закон предусматривает требования по прозрачности для поставщиков и пользователей определенных систем ИИ и моделей GPAI, включая

(1) поставщиков систем ИИ и GPAI, генерирующих синтетический аудио-, графический, видео- или текстовый контент,

(2) разработчиков систем распознавания эмоций или биометрической категоризации,

(3) разработчиков систем ИИ, которые генерируют или манипулируют изображениями, аудио- или видеоконтентом, составляющими подделку, и

(4) разработчиков систем, которые генерируют или манипулируют текстом, опубликованным с целью подделки или для информирования общественности по вопросам, представляющим общественный интерес. Закон налагает дополнительные обязательства по обеспечению прозрачности на лиц, внедряющих определенные системы ИИ высокого риска. В некоторых случаях контент должен быть помечен машиночитаемым способом, чтобы его можно было идентифицировать как искусственно созданный или подвергнутый манипуляциям. Закон об ИИ предусматривает исключения при некоторых обстоятельствах, в т. ч. когда система ИИ используется в художественных, сатирических, творческих или аналогичных целях.

Быстрорастущие модели ИИ общего назначения (GPAI) также должны будут соответствовать обязательствам по прозрачности и правилам ЕС об авторском праве, в то время как к наиболее мощным моделям будут предъявляться дополнительные требования безопасности. Учитывая возрастающие трудности с распознаванием искусственных или подделанных аудиовизуальных носителей («фейков») в Интернете, такой контент должен иметь четкую маркировку.

В Законе установлены и другие ограничения: согласованы гарантии в отношении ИИ общего назначения; предусмотрено ограничение использования систем биометрической идентификации правоохранительными органами; установлен запрет на социальный скоринг и ИИ, используемый для манипулирования уязвимостями пользователей или эксплуатации их уязвимостей; предусмотрено право потребителей подавать жалобы и получать содержательные объяснения.

Штрафы за нарушение Закона варьируются от 35 млн евро (7% от годового оборота) до 7,5 млн евро (1,5% от оборота).

Закон вступит в силу через 20 дней после его публикации в Официальном журнале ЕС и, как правило, начнет применяться к организациям через 2 года после

его вступления в силу, за некоторыми исключениями: запреты на определенные методы использования ИИ вступят в силу через 6 месяцев, правила в отношении моделей GPAI вступят в силу через 12 месяцев (за исключением моделей GPAI, которые были размещены на рынке до этой даты; они вступят в силу еще через 24 месяца), а также правила, применимые к ИИ высокого риска, включенному в Приложение II: системы вступят в силу через 36 месяцев.

Признавая потенциальную угрозы правам граждан, создаваемых определенными приложениями ИИ, **Закон запрещает**:

- системы биометрической категоризации, использующие конфиденциальные характеристики (например, политические, религиозные, философские убеждения, сексуальная ориентация, раса);
- нецелевое извлечение изображений лиц из Интернета или видеозаписей с камер видеонаблюдения для создания баз данных распознавания лиц;
- распознавание эмоций на рабочем месте и в учебных заведениях;
- социальный рейтинг, основанный на социальном поведении или личных характеристиках;
- системы ИИ, которые манипулируют поведением людей в обход их свободной воли;
- системы ИИ, использующие уязвимости людей (из-за их возраста, инвалидности, социального или экономического положения).

Строгие ограничения налагаются в отношении *использования систем биометрической идентификации* (RBI) в общедоступных местах в целях обеспечения правопорядка при условии предварительного разрешения суда и для строго определенных списков преступлений. Они должны использоваться строго при целенаправленном поиске лица, осужденного или подозреваемого в совершении серьезного преступления. RBI в режиме реального времени должны соответствовать строгим условиям, и их использование будет ограничено по времени и местоположению в следующих целях:

- целенаправленные поиски жертв (похищение, торговля людьми, сексуальная эксплуатация),
- предотвращение конкретной и существующей террористической угрозы или
- локализация или идентификация лица, подозреваемого в совершении одного из конкретных преступлений, упомянутых в постановлении (например, терроризм, торговля людьми, сексуальная эксплуатация, убийства, похищения людей, изнасилования, вооруженное ограбление, участие в преступной организации, экологические преступления).

Закон требует оценки воздействия на основные права, прежде чем система ИИ высокого риска будет выведена на рынок. Кроме того, для систем, которые (i) взаимодействуют с людьми, (ii) используются для обнаружения эмоций или определения связи с (социальными) категориями на основе биометрических данных или (iii) генерируют контент или манипулируют им («дипфейк»), обязательно информирование пользователей об их автоматизированном характере, например, для информирования физических лиц, когда они подвергаются воздействию системы распознавания эмоций, или для эффективного нанесения водяных знаков на контент, созданный ИИ, в соответствии с техническими стандартами. Граждане будут иметь право подавать жалобы на системы ИИ и получать разъяснения по поводу решений, полученных при помощи систем ИИ высокого риска, которые затрагивают их права.

С учетом широкого круга задач, которые могут выполнять системы ИИ, и быстрого расширения его возможностей Законом предусмотрено, что системы ИИ общего назначения (GPAI) и модели GPAI, на которых они основаны, должны будут соответствовать требованиям прозрачности, первоначально предложенным парламентом. Они включают составление технической документации, соблюдение законодательства ЕС об авторском праве и распространение подробных сведений о контенте, используемом для обучения.

Для высокоэффективных моделей GPAI с системным риском участникам парламентских переговоров удалось добиться более строгих обязательств. Если эти модели соответствуют определенным критериям, они должны будут проводить оценку моделей, оценивать системные риски и снижать их, проводить состязательное тестирование, сообщать Комиссии о серьезных инцидентах, обеспечивать кибербезопасность и сообщать об их энергоэффективности. Депутаты Европарламента также настаивали на том, что до публикации гармонизированных стандартов ЕС GPAI с системным риском могут полагаться на кодексы практики для соблюдения регламента.

Кроме того, Закон включает список запрещенных практик для тех систем ИИ, использование которых считается неприемлемым как противоречащее ценностям ЕС, таких как когнитивно-поведенческие манипуляции или обман, использование уязвимостей, нецелевое извлечение изображений лиц из Интернета или видеозаписей с камер видеонаблюдения, классификация социального поведения, социальный рейтинг, биометрическая категоризация для вывода конфиденциальных данных, таких как сексуальная ориентация или религиозные убеждения, и некоторые случаи превентивного полицейского воздействия на отдельных лиц.

Законом создан специальный надзорный орган ЕС — Офис искусственного интеллекта при Еврокомиссии, который будет осуществлять надзор за самыми передовыми моделями ИИ, способствовать внедрению стандартов и практик тестирования, а также обеспечивать соблюдение общих правил во всех государствах-членах. Научная группа независимых экспертов будет консультировать Офис искусственного интеллекта по моделям GPAI. Совет по искусственному интеллекту будет действовать как координационная платформа и консультативный орган при Комиссии, будет также

создан консультативный форум для заинтересованных сторон, таких как представители промышленности, малого и среднего бизнеса, стартапы, гражданское общество и научные круги, для предоставления технической экспертизы Совету по искусственному интеллекту.

Поставщики автономных систем ИИ высокого риска и определенные пользователи систем ИИ высокого риска, которые являются государственными организациями, должны зарегистрироваться в базе данных ЕС по системам ИИ высокого риска.

Поставщики систем ИИ обязаны также соблюдать обязательства по мониторингу и отчетности в отношении пострыночного мониторинга и отчетности, а также расследования инцидентов и неисправностей, связанных с ИИ.

Заключение

С быстрым распространением технологий ИИ он стал значительной силой, которая меняет методы работы предприятий и взаимодействия людей с машинами в различных производственных сферах, включая:

- *автоматизацию* повторяющихся задач, повышение эффективности и уменьшение количества человеческих ошибок;
- быстрый *анализ обширных наборов данных* ИИ для принятия решений на их основе;
- *персонализация* и адаптация опыта пользователей с представлением рекомендаций по контенту или спросу на продукцию на основе индивидуальных предпочтений и потребительского поведения;
- *стимулирование инноваций*, способствуя разработке новых продуктов, услуг и иных решений, которые ранее были недостижимы, за счет автоматизации тестирования или генерации новых формул;
- *улучшение принятия управленческих решений* — ИИ может предоставлять информацию и прогнозы, помогающие лицам, принимающим решения, в различных областях, от диагностики здравоохранения до финансового прогнозирования;
- *повышение безопасности* на базе систем ИИ благодаря таким функциям, как автономное вождение и прогнозируемое техническое обслуживание и др.;
- *создание нового контента* — алгоритмы ИИ используются для создания контента в литературе, искусстве и других областях человеческой деятельности. Например, Creative Commons использует генеративный ИИ для создания контента способами, которые поддерживают открытый доступ к образованию и творческим разработкам.

Литература

1. Карцхия А.А., Макаренко Г.И. Правовые горизонты технологий искусственного интеллекта: национальный и международный аспект // Вопросы кибербезопасности. 2024. № 1. С. 2—14. DOI: 10/21681/2311-3456-2024-1-2-14 .
2. Мохов А.А. Демографическая безопасность и ее правовое обеспечение // Юрист. 2023. № 6. С. 62—67.
3. Amatova N.E., Social consequences of the implementation of NBIC-technologies: risks and expectations. Univ. Soc. Sci. 9 (8) (2014). URL: http://7universum.com/en/social/archive/item/1549 (accessed 22 Jan 2020).
4. S. Klaus, C. Jung, 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-a020-5b4063662b35/sw_newsletter_october_i_english.pdf
5. Haskins A., Arora S., Nilawar U. Impact of Artificial Intelligence on Indian Real Estate: Transformation Ahead. Collier's Radar Property Research (India). 05.10.2017. 13 p. P. 4.
6. Capabilities and risks from frontier AI, AI Safety Summit, 2023. URL: https://assets.publishing.service.gov.uk/media/65395abae6c968000daa9b25/frontier-ai-capabilities-risks-report.pdf
7. Frontier AI Regulation: Managing Emerging Risks to Public Safety, November 7, 2023. URL: https://arxiv.org/abs/2307.03718
8. The Paradox of Artificial Intelligence in the Legal Industry: Both Treasure Trove and Trojan Horse? The Perils of Deepfakes, Wolters Kluwer, 2021. URL: http://arbitrationblog.kluwerarbitration.com
9. Марков А.С. Важная веха в безопасности открытого программного обеспечения // Вопросы кибербезопасности. 2023. № 1 (53). С. 2—12.
10. Карцхия А.А. LegalTech как основа цифровой правовой экосистемы / LegalTech в сфере предпринимательской деятельности : монография (отв. ред. И.В. Ершова, О.В. Сушкова). М. : Проспект, 2023. С. 25—33.
11. Карцхия А.А., Макаренко Г.И., Макаренко Д.Г. Правовые перспективы технологий искусственного интеллекта // Безопасные информационные технологии : сборник трудов Двенадцатой международной научно-технической конференции МВТУ им Н. Э. Баумана. 2023. С. 154—161.
12. Крутских А.В., Зиновьева Е.С. Международная информационная безопасность: подходы России. М. : МГИМО МИД России, 2021. С. 6.

LEGAL PROBLEMS IN USING ARTIFICIAL INTELLIGENCE IN RUSSIA

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

Grigory Makarenko, *Senior Researcher at the Scientific Centre for Legal Information under the Ministry of Justice of the Russian Federation, Moscow, Russian Federation.*
E-mail: t7920518@yandex.ru

Keywords: *neural networks, machine learning, artificial intelligence security, legal basis.*

Abstract

Purpose of the work: showing the legal problems in the development and implementation of artificial intelligence (AI) in Russian reality.

Study findings. A review of the topic for Russia, USA, and China was carried out. Although Russia is at the 10th place in the world in terms of using AI, its implementation is progressing rapidly. The authors wanted to show (and to warn) that the implementation of what is now called AI was under development already in the USSR. Back in the 1970, one of the authors of this paper set up a computer-aided design laboratory for automated design of 13-layer printed circuit boards for on-board computers (used in aerospace systems). By 1980 there were already hundreds of units in the USSR, in different fields of technology and industry, engaged in the automation of design and management. The development of automation in Russia stopped due to the stop of the development of industry in the country: passenger aircraft industry, machine tool industry, instrumentation engineering were nearly completely destroyed, and it is only in the recent years that the country came to its senses and started to discuss the development of its industry. However, as we can see from the example of aircraft industry, even passenger aircraft that were tested long ago and already produced before still can't reach the production stage.

There stand not only artificial obstacles such as negligent officials in the way of implementation of AI but also objective circumstances such as a lack of a legal basis.

Practical importance: this paper is a follow-up to a paper by the authors "Legal horizons of artificial intelligence technologies: national and international aspects" (the Cybersecurity Issues journal, No. 1 for 2024) and may be useful in developing the said legal basis.

References

1. Kartskhiia A.A., Makarenko G.I. Pravovye gorizonty tekhnologii iskusstvennogo intellekta: natsional'nyi i mezhdunarodnyi aspekt. Voprosy kiberbezopasnosti, 2024, No. 1, pp. 2–14. DOI: 10/21681/2311-3456-2024-1-2-14 .
2. Mokhov A.A. Demograficheskaia bezopasnost' i ee pravovoe obespechenie. Iurist, 2023, No. 6, pp. 62–67.
3. Amatova N.E., Social consequences of the implementation of NBIC-technologies: risks and expectations. Univ. Soc. Sci. 9 (8) (2014). URL: http://7universum.com/en/social/archive/item/1549 (accessed 22 Jan 2020).
4. S. Klaus, C. Jung, 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-a020-5b4063662b35/sw_newsletter_october_i_english.pdf
5. Haskins A., Arora S., Nilawar U. Impact of Artificial Intelligence on Indian Real Estate: Transformation Ahead. Collier's Radar Property Research (India). 05.10.2017. 13 p. P. 4.
6. Capabilities and risks from frontier AI, AI Safety Summit, 2023. URL: https://assets.publishing.service.gov.uk/media/65395abae6c968000daa9b25/frontier-ai-capabilities-risks-report.pdf
7. Frontier AI Regulation: Managing Emerging Risks to Public Safety, November 7, 2023. URL: https://arxiv.org/abs/2307.03718
8. The Paradox of Artificial Intelligence in the Legal Industry: Both Treasure Trove and Trojan Horse? The Perils of Deepfakes, Wolters Kluwer, 2021. URL: http://arbitrationblog.kluwerarbitration.com

9. Markov A.S. Vazhnaia vekha v bezopasnosti otkrytogo programmnogo obespecheniia. Voprosy kiberbezopasnosti, 2023, No. 1 (53), pp. 2–12.
10. Kartskhiia A.A. LegalTech kak osnova tsifrovoi pravovoi ekosistemy. LegalTech v sfere predprinimatel'skoi deiatel'nosti : monografiia (otv. red. I.V. Ershova, O.V. Sushkova). M. : Prospekt, 2023, pp. 25–33.
11. Kartskhiia A.A., Makarenko G.I., Makarenko D.G. Pravovye perspektivy tekhnologii iskusstvennogo intellekta. Bezopasnye informatsionnye tekhnologii : sbornik trudov Dvenadtsatoi mezhdunarodnoi nauchno-tekhnicheskoi konferentsii MVTU im N. E. Baumana, 2023, pp. 154–161.
12. Krutskikh A.V., Zinov'eva E.S. Mezhdunarodnaia informatsionnaia bezopasnost': podkhody Rossii. M. : MGIMO MID Rossii, 2021, p. 6.