ПРАВОВЫЕ ГОРИЗОНТЫ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: НАЦИОНАЛЬНЫЙ И МЕЖДУНАРОДНЫЙ АСПЕКТ.

Карцхия А. А.[1], Макаренко Г. И.[2]

DOI: 10.21681/2311-3456-2024-1-2-14

Цель исследования *– анализ факторов стремительного развития искусственного интеллекта и его потенциала, исследование новых моделей ИИ для повышения производительности труда, поощрения инноваций и формирования новых предпринимательских структур, а также решения социальных проблем в здравоохранении, образовании, разрешении климатического кризиса и достижении целей устойчивого развития.*

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

Результат: *в исследование показано, что внедрение программ искусственного интеллекта вместе с преимуществами создает трудно прогнозируемые угрозы и риски, имеющие трансграничный характер. С целью смягчения потенциальных опасностей, для обеспечения контролируемости и устойчивости технологий ИИ на основе концепции доверенного (ответственного) искусственного интеллекта необходимо утверждение руководящих принципов по искусственному интеллекту и создания универсального кодекса поведения разработчиков ИИ, которые совместно могут создать базу для единых основ правового регулирования в рамках национального законодательства каждой страны на основе принципов защиты прав человека, конфиденциальности и защиты данных, а также прозрачности и объяснимости, справедливости, подотчетности и безопасности ИИ, надлежащего контроля со стороны человека и этических норм создания и применения ИИ.*

Новизна исследования *заключается в том, что на основе риск ориентированного подхода предлагается концептуальная оценка полезности эффективности, устойчивости и безопасности технологий и моделей ИИ, а также установления его правового статуса, в том числе, для защиты человека от неконтролируемого влияния ИИ и неизменности гарантий конституционных прав и свобод человека.*

Ключевые слова: *международная кибербезопасность, доверенный интеллект, нейронные сети, машинное обучение, безопасность искусственного интеллекта, технологический суверенитет, угрозы и риски технологий, устойчивое развитие.*

LEGAL HORIZONS OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES: NATIONAL AND INTERNATIONAL ASPECTS.

Kartskhia A. A.[3], Makarenko G. I.[4]

The purpose *of the study is to analyze the factors of rapid development of artificial intelligence and its potential, to study new AI models to increase labor productivity, encourage innovation and the formation of new business structures, as well as solve social problems in healthcare, education, solving the climate crisis and achieving sustainable development goals.*

Research methods: *comparative legal method and methods of analysis and synthesis in the process of historical genesis of artificial intelligence, application of the risk-oriented method of AI assessment.*

Result: *The study shows that the introduction of artificial intelligence programs, together with the benefits, creates hard-to-predict threats and risks of a cross-border nature. In order to mitigate potential dangers, in order to ensure the controllability and sustainability of AI technologies based on the concept of trusted (responsible) artificial intelligence, it is necessary to approve guidelines on artificial intelligence and create a universal code Together, AI can create a framework for a common regulatory framework in each country's national legislation based on the principles of human rights, privacy and data protection, as well as transparency and explainability, fairness, accountability and safety of AI, appropriate human control, and ethical standards for the creation and use of AI.*

1 Карцхия Александр Амиранович, доктор юридических наук, профессор РГУ нефти и газа (НИУ) имени И. М. Губкина, Москва. Россия. E-mail: arhz50@mail.ru
2 Макаренко Григорий Иванович, старший научный сотрудник НЦПИ при Минюсте РФ, Москва, Россия. E mail: t7920518@yandex.com

3 Alexander A. Kartskhia, Doctor of Law, Professor, Gubkin Russian State University of Oil and Gas, Moscow. Russia. E-mail: arhz50@mail.ru
4 Grigory I. Makarenko, Senior Researcher, National Center for Strategic Studies under the Ministry of Justice of the Russian Federation, Moscow, Russia. E-mail: t7920518@yandex.com

The novelty of the research *lies in the fact that, on the basis of a risk-oriented approach, a conceptual assessment of the usefulness of the efficiency, sustainability and safety of AI technologies and models, as well as the establishment of its legal status, including for the protection of a person from the uncontrolled influence of AI and the immutability of guarantees of constitutional human rights and freedoms, is proposed.*

Keywords: *international cybersecurity, trusted intelligence, neural networks, machine learning, artificial intelligence security, technological sovereignty, threats and risks of technologies, sustainable development.*

Введение. Искусственный интеллект в центре всеобщего внимания

Мир стоит на пороге нового ренессанса в науке и технике, основанного на всестороннем понимании структуры и поведения материи от наноразмерных величин до самой сложной из когда-либо открытых систем - человеческого мозга. При должном внимании к этическим вопросам и потребностям общества результатом может стать значительное улучшение человеческих способностей, новых отраслей промышленности и продуктов, результатов для общества и качества жизни. При этом, все более актуальными становятся вопросы правового регулирования этих сфер деятельности [1,2,3][5].

Как отмечается в Концепции внешней политики Российской Федерации[6], человечество переживает эпоху революционных перемен. Структурная перестройка мировой экономики, связанная с переходом на новую технологическую основу посредством внедрения технологий искусственного интеллекта, новейших информационно-коммуникационных, энергетических, биологических технологий и нанотехнологий, а также рост национального самосознания, культурно-цивилизационное разнообразие и другие объективные факторы ускоряют процессы перераспределения потенциала развития в пользу новых центров экономического роста и геополитического влияния.

Экспоненциальные улучшения технологий искусственного интеллекта и других передовых технологий в последнее время привели к взрывному росту интереса (научного, коммерческого, военного и др.) в искусственный интеллект и финансовых инвестиций в него.

В настоящее время особое внимание приковано к *генеративному искусственному интелекту*. Генеративный искусственный интеллект (*AGI*) – это тип искусственного интеллекта, который может создавать (генерировать) новый контент и идеи, включая разговоры, истории, изображения, видео и музыку. Как и любой искусственный интеллект, генеративный ИИ основан на моделях машинного обучения – очень больших моделях, предварительно обученных на **огромных** объемах данных и обычно называемых базовыми моделями (FM). Базовые модели (FM), обученные работе с огромными наборами данных, представляют собой крупные нейронные сети с глубоким обучением, которые изменили подход специалистов по работе с данными к машинному обучению (ML). Вместо того чтобы разрабатывать искусственный интеллект с нуля, специалисты по работе с данными используют базовую модель в качестве отправной точки для разработки моделей ML, позволяющих быстрее и экономичнее поддерживать новые сферы применения. Термин «*базовая модель*» был придуман исследователями для описания моделей ML, обученных на широком спектре обобщенных и немаркированных данных и способных выполнять широкий спектр общих задач, таких как понимание языка, генерирование текста и изображений и общение на естественном языке. Технологии искусственного интеллекта пытаются имитировать человеческий интеллект в таких нетрадиционных вычислительных задачах, как распознавание изображений, обработка естественного языка (NLP) и перевод. Генеративный искусственный интеллект является следующим шагом в разработке искусственного интеллекта[7].

Генеративный ИИ быстро вошел в общественный дискурс. Стремительный прогресс в области генеративного искусственного интеллекта обусловлен его ожидаемым потенциалом для повышения производительности, поощрение инноваций и предпринимательства и поиск решений глобальных проблем, а также в решении социальных проблем, таких как улучшение здравоохранения и помощь в разрешении климатического кризиса и достижении Целей устойчивого развития (ЦУР).

5 Roco, M.C., Bainbridge, W.S. (2003). Overview Converging Technologies for Improving Human Performance. In: Roco, M.C., Bainbridge, W.S. (eds) Converging Technologies for Improving Human Performance. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-0359-8_1

6 Указ Президента РФ от 31.03.2023 N 229 «Об утверждении Концепции внешней политики Российской Федерации»// Собрание законодательства РФ, 03.04.2023, N 14, ст. 2406

7 https://aw.club/global/ru/blog/ai/generative-ai-for-content-creation (дата обращения 11 января 2024 г.)

Современные понятия искусственного интеллекта различаются в своих определениях. Некоторые фокусируются на способности программы делать любые прогнозы, рекомендации или решения. Например, NIST (*NIST Artificial Intelligence Risk Management Framework*) определяет «систему ИИ» в широком смысле как любую «машинную систему, которая может для заданного набора целей генерировать результаты, такие как прогнозы, рекомендации или решения, влияющие на реальную или виртуальную среду»[8]. Другие определения дают более узкое определение термина для обозначения программ, которые либо приближают человеческое мышление, либо заменяют его, т.е. программы, которые способны приближаться к человеческим, интеллектуальным способностям.

Искусственный интеллект рассматривается в российской Национальной стратегии развития искусственного интеллекта и российском законодательстве[9] как комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека. Такой комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру (в том числе информационные системы, информационно-телекоммуникационные сети, иные технические средства обработки информации), программное обеспечение (в т.ч. в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений.

При этом следует учесть, что технологии ИИ классифицируются в трех разных аспектах: методы, применяемые при создании ИИ (например, машинное обучение); функциональные приложения (например, обработка речи и компьютерное зрение); и области применения этих технологий (например, связь, транспорт)[10].

Стоит отметить, что термин «искусственный интеллект», как отмечают эксперты [4,5][11], не относится к какой-либо конкретной технологии – скорее, это собирательный термин для множества технологий использования математико-статистических методов для моделирования когнитивных способностей. Технологии искусственного интеллекта работают на основе анализа большого объема неструктурированных данных (*Big Data*) по специально разработанному алгоритму для выявления определенных закономерности 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 во взаимосвязанном 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*). Качественный уровень технологии ИИ зависит от его архитектуры, обучения и качества обучающих данных, поскольку структура нейронной сети, ее настройки должны быть адаптированы к конкретной цели, на которую нацелен ИИ (например, распознавание речи или изображений, генерация текста и т.д.). В идеале приложение искусственного интеллекта должно быть способно идентифицировать в большом объеме данных (например, в потоке данных камеры наблюдения) тип шаблона, для которого оно было обучено (например, лица, номерные знаки и т. д.), за очень короткое время. Фактический показатель успешности приложений искусственного интеллекта во многом зависит от структуры нейронной сети, способа ее обучения и качества используемых обучающих данных.

В связи с этим, важно установить **современное понимание искусственного интеллекта**. Обычно, *генеративный (общий) искусственный интеллект* (*AI*)

8 Shukla Shubhendu et al, Applicability of Artificial Intelligence in Different Fields of Life, 1 Int'l J. of Scientific Engineering and Research 1 at 28 (Sept. 2013).

9 Указ Президента РФ от 10.10.2019 N 490 «О развитии искусственного интеллекта в Российской Федерации // Собрание законодательства РФ, 14.10.2019, N 41, ст. 5700; Федеральный закон от 24.04.2020 N 123-ФЗ// Собрание законодательства РФ, 27.04.2020, N 17, ст. 2701

10 WIPO Technology Trends 2019, Artificial Intelligence. URL: https://www.theblockchaintest.com/uploads/resources/WIPO%20-%20Technology%20Trends%202019-Artificial%20Intelligence%20-%202019.pdf

11 Russell S., Norvig P. Artificial Intelligence: A Modern Approach. Third Edition. Boston: Prentice Hall, 2010. xviii; 1132 p. P. 1–2;
Rissland E.L. Artificial Intelligence and Law: Stepping Stones to a Model of Legal Reasoning // The Yale Law Journal. 1990. Vol. 99. N 8. P. 1957-1981. P. 1958-1959.

на основе машинного обучения использует нейронные сети и другие алгоритмы для создания новых данных или контента, похожих на исходные данные. Этот подход отличается от дескриптивного AI, который анализирует и классифицирует данные, но не создает новых данных. Генеративный AI может иметь огромное значение для различных отраслей, таких как медиа, искусство, развлечения, реклама и образование. Однако, он также может вызывать определенные угрозы в связи с нарушением авторских прав, распространением ложной или дискриминационной информации и потери контроля над созданным контентом.

Дескриптивный искусственный интеллект (AI) на основе машинного обучения используется для анализа, классификации и предсказания на основе необработанных данных и определяет структуру, зависимости и тенденции данных, не создавая новых данных. Дескриптивный AI может быть использован для различных целей, таких как: (a) классификация, т.е. разделение данных на группы на основе их характеристик или признаков (классификация электрокардиограмм (ЭКГ) на нормальные и аномальные, диагностика заболеваний и др.); (b) регрессия, т.е. предсказание неизвестных значений на основе известных данных (прогноз погоды, биржевых котировок и др.); (c) кластеризация, т.е. разделение данных на группы на основе схожести между элементами (моделирование бизнес-процессов и др.); (d) анализ тенденций, т.е. определение тенденций и зависимостей в данных для получения информации о будущих событиях или изменениях. Дескриптивный AI является основой для многих современных технологий, таких как рекомендательные системы, системами автоматической обработки звука и изображений, системами контроля качества и системами управления рисками. Хотя дескриптивный AI не создает новых данных, он может предоставить важную информацию и знания, которые могут быть использованы для принятия решений, планирования и стратегического планирования.

Вместе с тем, уже разработан новый вид ИИ – ***саморазвивающийся*** *искусственный интеллект*. Как заявили ученые Массачусетского технологического института и Калифорнийского университета (Fox News)[12], возможно создание подсистем ИИ без помощи человека. Более крупные модели ИИ, подобные тем, которые используют ChatGPT, на основе «родительского» алгоритма могут создавать меньшие, специфичные приложения искусственного интеллекта, которые можно применять, например, для усовершенствования слуховых аппаратов, мониторинга нефтепроводов или отслеживания исчезающих видов живой природы.

Ведущие страны в сфере разработки искусственного интеллекта, опираясь на активную поддержку государства, стремительными темпами развивают национальные технологии ИИ. После разработок Deep Mind и запуска в ноябре 2022 американской Open AI *ChatGPT* последовали публичные старты аналогичных технологий на основе LLM в других странах. В ноябре 2023 года в Абу-Даби (ОАЭ) была запущена поддерживаемая государством компания по искусственному интеллекту AI71 для коммерциализации модели ИИ LLM Falcon. В декабре того же года объявлено о масштабном финансировании французского ИИ Mistral. В Индии создаются национальные модели LLM Krutrim и Sarvam. Государства и частные компании в США, КНР, Великобритании, Франции, Германии, Индии, Саудовской Аравии и Объединенных Арабских Эмиратах (ОАЭ) масштабно финансируют разработки ИИ и развивают национальные производства графических редакторов (GPU-чипов) и других элементов, необходимых для создания ИИ[13]. В России созданы аналогичные разработки и достижения на базе нейросети ПАО Сбербанк России (RuGPT-3).

Международно-правовые аспекты статуса искусственного интеллекта

На первом международном саммите по безопасности искусственного интеллекта, прошедшем 1 ноября 2023г. в Великобритании (Россия не принимала в нем участие), страны-участники, включая США, Европейский Союз, Великобританию, Францию, Германию, Италию, КНР, Австралию Индию, Бразилию, Японию, Королевство Саудовской Аравии, Объединенные Арабские Эмираты, Нигерию и Кению, а также компании-лидеры IT отрасли (Amazon Web Services, Anthropic, Google, Google DeepMind, Inflection AI, Microsoft, Mistral AI, Open AI и xAI) подписали Декларацию по вопросам безопасности искусственного интеллекта (*The Bletchley Declaration on AI safety*)[14] (далее – Декларация), в которой устанавливается, что искусственный интеллект открывает огромные глобальные возможности, обладая потенциалом трансформировать мир и повышать благосостояние людей, и потому, для всеобщего блага искусственный интеллект должен проектироваться, разрабатываться, развертываться и использоваться безопасным образом.

12 Ученые заявили о возможности ИИ воспроизводиться без участия человека, 17 декабря 2023. URL: https://vfokuse.mail.ru/article/uchenye-zayavili-o-vozmozhnosti-ii-vosproizvoditsya-bez-uchastiya-cheloveka-59040575/ (дата обращения 11.01.2024 г.)

13 Welcome to the era of AI nationalism. The Economist, January 1st, 2024. URL: https://www.economist.com/ business/2024/01/01/welcome-to-the-era-of-ai-nationalism?utm_content=article-link-2&etear=nl_today_2&utm_campaign=a.the-economist-today&utm_medium=email.internal-newsletter.np&utm_source=salesforce-marketing-cloud&utm_term=1/1/2024&utm_id=1840347

14 https://www.gov.uk/government/news/countries-agree-to-safe-and-responsible-development-of-frontier-ai-in-landmark-bletchley-declaration

Подтверждается необходимость безопасного развития искусственного интеллекта и использования его преобразующих возможностей во благо всех, инклюзивным образом во всем мире, включая сферу здравоохранения и образования, продовольственной безопасности, науки, чистой энергетики, биоразнообразия и климата, а также для реализации прав человека и активизации усилий по достижению Целей устойчивого развития ООН. Однако огромные возможности ИИ сопряжены с рисками, которые могут угрожать глобальной стабильности.

В Декларации использован термин «*Frontier AI*» (*передовой, новаторский ИИ*), который представляет собой высокоэффективные модели ИИ общего назначения, которые могут выполнять широкий спектр задач и соответствовать или превосходить возможности, присутствующие в самых продвинутых моделях на сегодняшний день. В первую очередь это относится к большим языковым моделям (LLM), лежащим в основе ChatGPT,4 Claude,5 и Bard.6. Ведущие компании в области ИИ, такие как Open AI, DeepMind и Anthropic, разрабатывают большие языковые модели (LMS), такие как GPT-4, в два этапа: предварительное обучение и тонкая настройка. На предварительном этапе обучения LLM «читает» миллионы или миллиарды текстовых документов, обучаясь выстраивать слова. Во время тонкой настройки предварительно обученный ИИ дополнительно обучается на тщательно отобранных наборах данных, которые ориентированы на более специализированные задачи или структурированы таким образом, чтобы направлять поведение модели в соответствии с ценностями разработчика и ожиданиями пользователей. Модели Frontier AI все чаще становятся мультимодальными: в дополнение к тексту они могут генерировать и обрабатывать другие типы данных (изображения, видео и звук). Ключевыми входными данными для разработки являются вычислительные ресурсы для обучения и запуска модели, данные, на основе которых она может учиться, алгоритмы, определяющие этот процесс обучения, а также таланты и опыт, которые обеспечивают все это [6].

Также отмечено, что ИИ также создает **значительные риски**, что обусловливает необходимость решения вопросов защиты прав человека, прозрачности и объяснимости, справедливости, подотчетности, правового регулирования и безопасности, надлежащего контроля со стороны человека, этики, смягчения предвзятости, конфиденциальности и защиты данных. Отмечены потенциальные непредвиденные риски, связанные со способностью ИИ манипулировать контентом или генерировать вводящий в заблуждение контент. Особые риски безопасности возникают при использовании передового искусственного интеллекта, под которым понимаются те высокомощные модели искусственного интеллекта общего назначения, включая базовые модели, которые могут выполнять широкий спектр задач, а также соответствующие специфические узконаправленные модели ИИ, которые могут демонстрировать возможности, причиняющие вред, которые соответствуют или превосходят возможности, присутствующие в самых передовых моделях сегодняшнего дня. Существенные риски могут возникнуть из-за потенциального преднамеренного неправильного использования или непреднамеренных проблем контроля, связанных с согласованием с намерениями человека. Отчасти эти проблемы связаны с тем, что эти возможности не до конца поняты и поэтому их трудно предсказать. Особую обеспокоенность вызывают риски в сфере кибербезопасности и биотехнологий, а также усиливающиеся риски, связанные с дезинформацией. Существует потенциал для серьезного, даже катастрофического ущерба, преднамеренного или непреднамеренного, вытекающего из наиболее значительных возможностей моделей передового ИИ.

Многие риски, связанные с ИИ по своей природе интернациональный характер, и поэтому необходимо международное сотрудничество, чтобы обеспечить ориентированный на человека, заслуживающий доверия и ответственный искусственный интеллект, который безопасен и служит всеобщему благу. Сотрудничество могло бы включать в себя, где это уместно, классификацию рисков на основе национальных условий и применимых правовых рамок, а также разработку общих принципов и кодексов поведения в области ИИ.

Декларация содержит общее понимание возможностей и рисков, связанных с генеративным искусственным интеллектом (AGI), и понимание настоятельной необходимости осознания потенциальных рисков ИИ и коллективного управления ими посредством новых совместных глобальных усилий по обеспечению безопасной и ответственной разработки и внедрения передового ИИ. Страны-участницы согласились, что существенные риски могут возникнуть в результате потенциального преднамеренного неправильного использования или непреднамеренных проблем с контролем передового ИИ, при этом особую озабоченность вызывают риски кибербезопасности, биотехнологии и дезинформации. Среди основных рисков выделены такие, как предвзятость и нарушение конфиденциальности в применении ИИ. Особое вниманием уделено таким правовым аспектам, как нормативное регулирование передовых технологий, конфиденциальности и защиты данных, а также интеллектуальная собственность.

Другим важным международным событием последнего времени в сфере регулирования искусственного интеллекта стал организованный рядом западных стран так называемый Хиросимский процесс. Группа стран G7 30 октября 2923г. в г. Хиросима (Япония) приняли совместную Декларацию «*G7 Leaders' Statement on the Hiroshima AI Process*» [15], в составе которой приняты два основных документа: свод Международных руководящих принципов по искусственному интеллекту (*The International Guiding Principles on Artificial Intelligence*) и рекомендован Кодекс поведения для разработчиков искусственного интеллекта (*Code of Conduct for AI developers*) [16], содержащий набор правил, которым рекомендуется следовать разработчикам ИИ на добровольной основе для снижения рисков на протяжении всего жизненного цикла ИИ. Хиросимский процесс задуман с целью создания всеобъемлющей политической основы, способствующей разработке безопасных и заслуживающих доверия систем искусственного интеллекта и снижающей риски, возникающие, в частности, от генеративного искусственного интеллекта. Основными пятью рисками признаются: распространение дезинформации и манипулирование, нарушения интеллектуальной собственности, угрозы конфиденциальности, дискриминация и предвзятость, а также риски для безопасности. В Декларации отмечается, что решение задач управления рисками ИИ, исходя из общих принципов верховенства закона и демократических ценностей, требует формирования инклюзивного управления искусственным интеллектом на основе предложенных Международных руководящих принципов и Кодекса поведения для организаций, разрабатывающих передовые системы искусственного интеллекта. Предусматривается, что усилия в сфере ИИ в рамках Хиросимского процесса совместно с Глобальным партнерством по искусственному интеллекту (GPAI) и Организацией экономического сотрудничества и развития (ОЭСР) с участием многих заинтересованных участников, в т.ч. с правительствами, научными кругами, гражданским обществом и частными компаниями не только в странах G7, но и за ее пределами, включая развивающиеся страны и страны с формирующейся рыночной экономикой, будут способствовать созданию открытой и благоприятной среды, в которой безопасные и заслуживающие доверия системы искусственного интеллекта проектируются, разрабатываются, развертываются и используются для максимизации преимуществ технологии при одновременном снижении связанных с ней рисков, для общего блага во всем мире, с целью устранения цифрового разрыва и достижения цифровой инклюзивности.

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 искусственного интеллекта. 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*)[17], который, как ожидается, будет принят в начале 2024 года и установит юридически обязательные правила разработки и использования ИИ. Вполне вероятно, что этот Закон создаст образец, по которому страны ЕС будут стремиться моделировать свои собственные законодательные акты в области ИИ.

Принципы служат руководством для организаций, разрабатывающих базовые модели (генеративный) ИИ, и включают следующие 11 ключевых принципов, присущих так называемому «доверенному искусственному интеллекту»:

1. Принятие необходимых мер для выявления, оценки и снижения рисков на протяжении всего жизненного цикла ИИ от разработки до промышленного применения;
2. Выявление и устранение уязвимостей ИИ, включая инциденты и схемы неправильного использования или внедрения, в т.ч. при размещение на рынке;
3. Обнародование сведений о возможностях, ограничениях и областях надлежащего и ненадлежащего использования передовых систем ИИ для поддержания и обеспечения достаточной прозрачности и повышению подотчетности;
4. Применение ответственного обмена информацией и сообщениями об инцидентах среди организаций, разрабатывающих передовые системы ИИ, в т.ч. в промышленности, госуправлении, гражданском обществе и научном сообществе;
5. Разработка, внедрение и раскрытие политики управления ИИ и рисками, основанной на риск-ориентированном подходе, включая политику конфиденциальности и меры по смягчению последствий, в частности для организаций, разрабатывающих передовые системы ИИ;
6. Создание надежных средств контроля безопасности, включая физическую безопасность, кибербезопасность и защиту от внутренних угроз на протяжении всего жизненного цикла ИИ;

15 https://digital-strategy.ec.europa.eu/en/library/g7-leaders-statement-hiroshima-ai-process

16 G7 hiroshima process on generative artificial intelligence (AI): towards A G7 common understanding on generative AI, September 2023, OECD 2023 // http://www.oecd.org/termsandconditions

17 https://artificialintelligenceact.eu/

7. Разработка и внедрение надежных механизмов аутентификации контента и определения происхождения, позволяющие пользователям идентифицировать контент, созданный искусственным интеллектом;
8. Приоритетное внимание исследованиям, направленным на снижение социальных рисков и рисков безопасности ИИ, а также инвестициям в эффективные меры по их снижению;
9. Приоритетное внимание разработке передовых систем искусственного интеллекта для решения глобальных мировых проблем, включая, но не ограничиваясь проблемами климатического кризиса, глобального здравоохранения и образования;
10. Поощрение разработки и принятие международных технических стандартов;
11. Обеспечение мер по вводу данных и защите персональных данных и интеллектуальной собственности.

Единообразие подходов при определении ИИ, общие определения для ИИ на международном уровне и в разных секторах его применения способны обеспечить инклюзивный диалог, устраняя различия между юрисдикциями и способствуя междисциплинарным коммуникациям и сотрудничеству. Основой для таких усилий может служить Концепция Организации экономического сотрудничества и развития (ОЭСР) по классификации систем искусственного интеллекта[18].

ОЭСР обновила свое определение искусственного интеллекта, изложив его в Рекомендациях: «Система искусственного интеллекта – это машинная 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 искусственного интеллекта различаются по уровню автономии и адаптивности после развертывания.»

Концепция определяет, что ***модель искусственного интеллекта*** представляет собой вычислительное представление всей внешней среды системы искусственного интеллекта или ее части, охватывающее, например, процессы, объекты, идеи, людей и/или взаимодействия, которые происходят в этой среде.

Модели искусственного интеллекта используют данные и/или экспертные знания, предоставляемые людьми и/или автоматизированными инструментами, для представления, описания и взаимодействия с реальной или виртуальной средой. При этом, выделяется ИИ «в лабораторных условиях» (*AI «in the lab»*), что относится к концепции и разработке системы ИИ до ее развертывания. Это применимо к данным и входным данным (например, для квалификации данных), модели искусственного интеллекта (например, для обучения исходной модели) и измерениям задачи и выходных данных (например, для задачи персонализации) фреймворка. Это особенно актуально для подходов и требований к управлению рисками *ex ante*. ИИ «в полевых условиях» (*AI «in the field»*) относится к использованию и эволюции системы ИИ после развертывания и применимо ко всем измерениям. Это относится к подходам и требованиям к управлению рисками *ex post*.

Основные характеристики включают технический тип, способ построения модели (с использованием экспертных знаний, машинного обучения или того и другого) и способ использования модели (для каких целей и с использованием каких показателей эффективности). Важно понимание *Жизненного цикла систем искусственного интеллекта*, который включает следующие этапы: (*a*) проектирование, данные и модели, представляющие контекстно-зависимую последовательность, охватывающую планирование и проектирование, сбор и обработку данных, а также построение моделей; (*b*) верификацию и валидацию; (*c*) развертывание; (*d*) эксплуатацию и мониторинг. Эти этапы часто выполняются интерактивным образом и не обязательно являются последовательными. Решение о выводе системы искусственного интеллекта из эксплуатации может быть принято в любой момент на этапе эксплуатации и мониторинга.

В документах OECD[19] сформулированы принципы ответственного управления заслуживающим доверия ИИ, которые дополняют друг друга и должны рассматриваться как единое целое. К ним, в частности, относятся:

- ✓ инклюзивный рост, устойчивое развитие и благосостояние, подразумевающие участие в ответственном управлении заслуживающим доверия искусственного интеллекта в целях расширения человеческих возможностей и креативности, содействия интеграции населения, сокращение экономического, социального, гендерного и других видов неравенства и защита природной среды, тем самым стимулируя инклюзивный рост, устойчивое развитие и благосостояние;

18 OECD framework for the classification of ai systems, OECD 2022. https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htm

19 OECD (2023), Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449; OECD (2020), Digitalisation and Responsible Business Conduct: Stocktaking of policies and initiatives. https://www.oecd.org/daf/inv/mne/ publicationsdocuments/reports/2/

- ✓ уважение верховенства закона и прав человека (свобода, достоинство и автономия, неприкосновенность частной жизни и защита данных, недискриминация и равенство, многообразие, честность, социальная справедливость и международно признанные трудовые права) и с этой целью внедрение соответствующих механизмов и гарантий, которые соответствуют контексту и соответствуют уровню техники;
- ✓ прозрачность и объяснимость, т.е. ответственное раскрытие значимой информации о системах ИИ, соответствующую контексту и согласующуюся с уровнем техники;
- ✓ надежность и защищенность ИИ на протяжении всего своего жизненного цикла, чтобы в условиях нормального использования, предсказуемого использования или неправильного использования или других неблагоприятных условий они функционировали надлежащим образом и не создавали необоснованного риска для безопасности;
- ✓ подотчетность, т.е. субъекты искусственного интеллекта должны нести ответственность за надлежащее функционирование систем искусственного интеллекта и за соблюдение вышеуказанных принципов, исходя из их ролей, контекста и в соответствии с уровнем техники.

Определение модели ИИ важно для государственной политики, поскольку ключевые свойства моделей ИИ – степень прозрачности и/или объяснимости, надежность и последствия для прав человека, неприкосновенности частной жизни и справедливости – зависит от типа модели, а также от процессов построения модели и логического вывода. Например, системы, использующие нейронные сети, часто рассматриваются как потенциально способные обеспечить сравнительно более высокую точность, но менее объяснимые, чем системы других типов. Объяснимость часто связана со сложностью системы; чем сложнее модель, тем труднее ее объяснить. Степень, в которой модель эволюционирует в ответ на данные, имеет отношение к государственной политике и режимам защиты прав потребителей, особенно для систем искусственного интеллекта, которые могут извлекать уроки из итераций и эволюционировать с течением времени. Понимание того, как была разработана и/или поддерживается модель, является еще одним ключевым фактором при распределении ролей и обязанностей в рамках процессов управления рисками. Субъекты искусственного интеллекта в этом измерении включают разработчиков и моделистов, которые создают, применяют, проверяют и контролируют модели ИИ.

Закон ЕС об искусственном интеллекте классифицирует технологии искусственного интеллекта по трем категориям риска, которые определяют сферу его использования и применения. Во-первых, запрещены к применению технологии и системы ИИ, которые создают неприемлемый риск, такие как государственная система социального скоринга (используемая в КНР). Во-вторых, особые требования установлены законом к технологиям ИИ с высоким уровнем риска, которые могут, к примеру, использоваться как инструмент сканирования резюме в целях ранжирования кандидатов при приеме сотрудников на работу. Остальные модели ИИ, которые явно не запрещены или не перечислены как высокорискованные, в значительной степени остаются нерегулируемыми.

Международный саммит по безопасности ИИ и Хиросимский процесс стали новой вехой в формировании международно-правового регулирования сферы цифровых технологий и прежде всего – искусственного интеллекта как наиболее перспективной и комплексной системы.

Вместе с тем обозначилось отсутствие стремления к установлению консенсуса по вопросам регулирования ИИ на международном уровне, что проявилось в собрании лишь небольшой группы государств (хотя и являющихся лидерами в сфере развития ИИ) для решения исключительно актуального вопроса – создания и развития технологий и моделей генеративного ИИ. Очевидно, решение проблемы безопасности ИИ должно привлекать значительно большее число стран, несомненно заинтересованных в установлении единых правил развития и применения передовых систем искусственного интеллекта.

Выступая на международной конференции «Путешествие в мир искусственного интеллекта AI Journey» 24 ноября 2023 Президент Российской Федерации заявил, что «с внедрением искусственного интеллекта в науку, образование, здравоохранение, все сферы нашей жизни — человечество начинает новую главу своего существования». Указывая на роль искусственного интеллекта сегодня, он отметил, что значение искусственного интеллекта имеет колоссальное значение и о того, каких результатов достигнет страна в соперничестве с другими государствами в сфере искусственного интеллекта, будет зависеть суверенитет, безопасность и состоятельность России. Поставлена задача дальнейшего укоренного развития технологий ИИ для того, чтобы Россия стала одной из самых комфортных юрисдикций для развития искусственного интеллекта. Необходимо разработать на основе генеративного искусственного интеллекта большие отраслевые модели, предложить механизмы их практического внедрения в целях существенного повышения производительности труда, а значит

и заработные платы в ключевых отраслях отечественной экономики. Также предлагается создать глобальные приемлемые для всех правила использования ИИ[20].

Современное понимание ИИ и его безопасность

Проблема безопасности ИИ сформулирована и обсуждается сравнительно недавно, но ее столь широкое правовое оформление на международном уровне сделано впервые. Безопасность искусственного интеллекта представляет собой состояние защищенности от угроз для человека при использовании ИИ, во взаимодействии с ИИ и в системе социальной и биосферы человечества, где ИИ уже стал значимым фактором, оказывающим самостоятельное влияние на общественные отношения, на самого человека. ИИ переходит от стадии инструмента, созданного человеком, к самостоятельной операционной системе, существующей по особым правилам и законам, техническим стандартам, и все чаще на базе машинного обучения (ML) и нейронных сетей.

Как отмечалось в Декларации Хиросимского процесса (2023г.)[21], потенциальные преимущества генеративного ИИ сопряжены с определенными рисками. Способность генеративного ИИ усугублять проблемы дезинформации и манипулирования мнениями рассматривается как одна из основных угроз, исходящих от генеративного ИИ, наряду с рисками нарушения прав интеллектуальной собственности и неприкосновенности частной жизни. Ответственное использование генеративного ИИ, борьба с дезинформацией, защита прав интеллектуальной собственности и управление генеративным ИИ являются одними из главных приоритетов и требуют международного сотрудничества. Другие неотложные и важные вопросы включают конфиденциальность и управление данными, прозрачность, справедливость и предвзятость, права человека и фундаментальные права, безопасность и надежность систем искусственного интеллекта, а также влияние на функционирование демократии.

Экспертами в области ИИ обсуждается вопрос о том, могут ли типы моделей ИИ в конечном итоге привести к созданию искусственному общему интеллекту (AGI), стадии, на которой автономные машины могли бы обладать возможностями человеческого уровня в самых разнообразных вариантах использования. Из-за его потенциального широкого воздействия на общество потенциальные выгоды и риски AGI заслуживают внимания, равно как и потенциально неизбежные последствия узких генеративных систем искусственного интеллекта, которые могут быть столь же значительными, как и AGI.

Долгосрочные преимущества и риски генеративного искусственного интеллекта могут потребовать решений в более широком, системном масштабе, чем уже применяемые подходы к снижению рисков.

Модели генеративного ИИ обучаются на огромных объемах данных, которые включают данные, защищенные авторским правом, в основном без разрешения правообладателей. Продолжающаяся дискуссия о способах защиты интеллектуальной собственности, и в частности, авторских прав, заключается в том, могут ли сами по себе искусственно созданные ИИ результаты интеллектуальной деятельности быть защищены авторским правом или запатентованы, и если да, то кто будет являться правообладателем.

Особый взгляд на ИИ, как отмечено в аналитическом материале OECD[22], связан с распространяемым в цифровых пространствах контентом, где он используется для обучения генеративных моделей искусственного интеллекта, что приводит к порочному негативному циклу в качестве онлайн-информации. Возрастают риски автоматизированных и персонализированных кибератак, слежки и цензуры, чрезмерной зависимости от генерирующих систем, научной чистоплотности разработчиков и концентрации власти и ресурсов.

В долгосрочной перспективе новые формы возможного «асоциального» поведения ИИ предполагают дополнительные риски, включая повышенную активность, стремление к власти и разработку неизвестных подцелей, определяемых машинами для достижения основных целей, запрограммированных человеком, но которые могут не соответствовать человеческим ценностям и намерениям.

Тем не менее, генеративный ИИ быстро внедряется в ключевых секторах промышленности. Прогнозируется, что генеративный ИИ в состоянии создавать существенную экономическую ценность и социальное благополучие. Компании начали внедрять технологии для создания новых бизнес-возможностей, а стартапы конкурируют за венчурный капитал. Популярные на сегодняшний день варианты использования программ-приложений включают предварительную обработку данных, сжатие и классификацию изображений, медицинскую визуализацию, персонализацию и интуитивно понятные пользовательские интерфейсы.

Системы генеративного искусственного интеллекта (ИИ) создают новый контент в ответ на запросы, основанные на их обучающих данных. Распространение систем генеративного искусственного интеллекта

20 http://www.kremlin.ru/events/president/transcripts/72811

21 G7 hiroshima process on generative artificial intelligence (ai): towards a G7 common understanding on generative AI, September 2023, OECD 2023. http://www.oecd.org/termsandconditions

22 Initial policy considerations for generative artificial intelligence, oecd artificial intelligence papers, September,2023. URL: http://www.oecd.org/

высветили возможности искусственного интеллекта, включая, например, ChatGPT для текста или для изображений (Stable Diffusion), для аудио или видео (DeepVoice), а также мультимодельные системы, объединяющие несколько типов медиа или языковые модели ИИ.

Однако эти же технологии создают критические социальные и политические проблемы, которые выражаются в потенциальных изменениях на рынках труда, неопределенности в отношении прав интеллектуальной собственности; риски, связанные с возможностью злоупотреблений при создании дезинформации и манипулируемого контента, распространением ложной информации (*deep fake*). В итоге могут формироваться негативные социальные, политические и экономические последствия, включая дезинформацию по ключевым научным вопросам, создание стереотипов и дискриминации, искажение общественного дискурса, создание и распространение теорий заговора и другой дезинформации, влияние на политические выборы, искажение рыночной информации и даже подстрекательство к насилию. Это, тем не менее, не отрицает преобразующую природу генеративного искусственного интеллекта и значимости международных дискуссий о стремлении к инклюзивному и заслуживающему доверия искусственному интеллекту.

Вместе с тем генеративный искусственный интеллект значительно увеличивает масштабы и объем дезинформации. В 2022 году было обнаружено, что люди почти в 50% случаев неспособны отличить искусственный интеллект от новостей, созданных человеком. Это означает, что генеративный ИИ может усилить риски как дезинформации (непреднамеренного распространения ложной информации), так и преднамеренной дезинформации злоумышленниками. Передовые модели генеративного искусственного интеллекта обладают мультимодальными возможностями, которые могут усугубить эти риски, например, путем объединения текста с изображением, видео или даже голосами[23].

Еще одной тревожной особенностью технологий ИИ является их склонность к «галлюцинациям» (т.е. к получению неверных, но убедительных результатов), особенно когда ответ отсутствует в данных обучения. Это позволяет может создавать убедительную дезинформацию, разжигать ненависть или воспроизводить предубеждения. Риски также включают чрезмерное доверие и чрезмерную зависимость от модели ИИ, что приводит к зависимости, которая может помешать развитию навыков и даже привести к потере навыков (OpenAI).

Синтетический контент может быть особенно полезен в политике, науке и правоохранительных органах. Риски, связанные с моделями искусственного интеллекта, генерирующими текст в изображение, ясно показывают, насколько быстр технологический прогресс.

Многочисленные «фотографии» в Twitter и других онлайн-платформах изображали известных политических деятелей и глав государств, совершающих неожиданные действия, но при этом вызывали большое доверие, демонстрируя силу синтетических образов, особенно в поляризованных политических контекстах. Другой проблемой было манипулирование научными изображениями для создания ложной информации (*deep fake*), угрожающей доверию в исследовательских и научных сообществах, а также репутации науки в глазах широкой общественности. В качестве примеров можно привести использование синтетических изображений отрицателями изменения климата или распространение дезинформации о COVID-19[24].

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

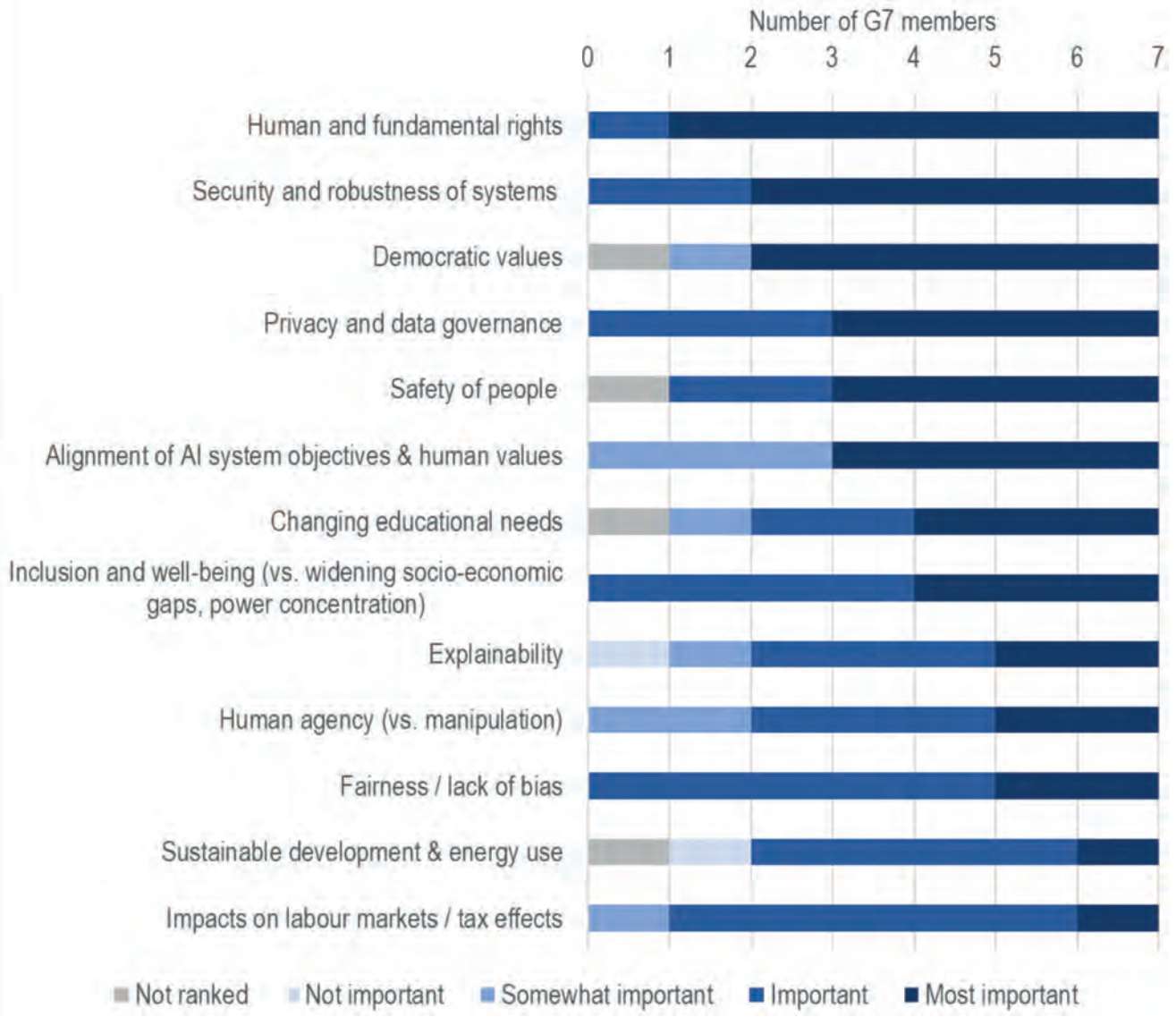
В последние годы все большее применение технологии искусственного интеллекта используют в целях упрощения, автоматизации и ускорения работы в различных областях [8], включая услуги по автоматизации заключения и мониторинга контрактов, программное обеспечение для электронного раскрытия информации, программные продукты для управления делами, агрегаторы информации, позволяющие принимать более обоснованные решения при выборе бизнес-решений или решений для прогнозирования судебных процессов. При этом, все острее встают вопросы безопасности и правомерности действий с учетом проблем имитации и ложных фактов, фальсифицированных и фальшивых фактов (Deep fake), создаваемых с помощью тех же технологий искусственного интеллекта. Предполагается, что решением этих проблем могло бы стать создание цифровой метки происхождения, т.е. встраивание цифровых «отпечатков пальцев» в соответствующие носители, с использованием возможностей технологии блокчейн. Мир искусства уже принял аналогичные процедуры, изобретательно и очень прибыльно используя так называемые NFT (*non-fungible tokens – невзаимозаменяемые токены*) для подтверждения

23 Initial policy considerations for generative artificial intelligence, oecd artificial intelligence papers, September,2023. URL: http://www.oecd.org/

24 Initial policy considerations for generative artificial intelligence, oecd artificial intelligence papers, September, 2023. URL: http://www.oecd.org/

Таблица 1

G7 Hiroshima process on artificial intelligence (AI)



уникальности произведений искусства, хранящихся в цифровом виде.

Как отмечают эксперты [9,10], в условиях обеспечения технологического суверенитета страны при усилении наступающей составляющей информационной безопасности идет совершенствование таксономий, развитие технологий и средств защиты, их адаптация под новые архитектуры (облачные, микросервисные и пр.), внедрение прорывных технологий (AI/ML, Big Data), а также формирование новых требований применения программного обеспечения с открытым исходным кодом, особенно в условиях новых угроз и рисков.

В этой связи целесообразно предусмотреть и согласовать принципы правового регулирования и юридически значимые этические принципы использования технологий искусственного интеллекта, внедрить стандартизированные правила его использования в правосудии и разрешении коммерческих споров, включая развитие современной цифровой криминалистики. Искусственный интеллект важен, в т.ч., с точки зрения понимания его функциональности и безопасности, а также соответствия принципу верховенства права и подконтрольности решений ИИ человеку [10].

Передовые высокоэффективные базовые модели ИИ могут обладать опасными возможностями, достаточными для создания серьезных рисков для общественной безопасности и регулирования. Для решения этих проблем необходимы по крайней мере три составных элемента регулирования моделей «*Frontier AI*»: (a) установление стандартов и соответствующих требований к разработчикам моделей ИИ; (b) установление требований и регистрации и отчетности для обеспечения регулирующим органам наглядность процессов разработки модели ИИ,

и (с) механизмы для обеспечения соответствия стандартам безопасности при разработке и внедрении моделей *Frontier AI*. Саморегулирование отрасли является важным первым шагом. Однако для создания стандартов безопасности и обеспечения их соответствия необходимо предоставить надзорным органам полномочий по обеспечению контроля соблюдения стандартов и режим лицензирования моделей *Frontier AI*.

Первоначальный набор стандартов безопасности ИИ включает: проведение оценок рисков перед развертыванием; внешний контроль поведения модели; использование оценок рисков для обоснования решений о развертывании, а также мониторинг и реагирование на новую информацию о возможностях модели и ее использовании после развертывания. *Frontier AI* способен выполнять широкий спектр задач и дополняется инструментами для расширения своих возможностей. Прогресс в течение следующих нескольких лет может быть быстрым и неожиданным в определенных отношениях. Вполне возможно, что в недалеком будущем могут быть разработаны продвинутые агенты ИИ общего назначения, но существует несколько глубоких, нерешенных сквозных технических и социальных факторов риска развития ИИ.[25]

Вопросы обеспечения международной информационной и кибербезопасности безопасности стали особенно актуальны как в практическом плане, так и в науке международного права в условиях нарастающих вызовов и угроз, связанных с использованием современных технологий в т. ч. против суверенитета государств, осуществления в глобальном информационном пространстве действий, препятствующих поддержанию международной безопасности и стабильности.

В связи с этим, в соответствии с Концепцией внешней политики Российской Федерации приоритетами России в целях обеспечения международной информационной безопасности, противодействия угрозам в ее отношении, укрепления российского суверенитета в глобальном информационном являются:

- ✓ укрепление и совершенствование международно-правового режима предотвращения и разрешения межгосударственных конфликтов и регулирования деятельности в глобальном информационном пространстве;
- ✓ формирование и совершенствование международно-правовых основ противодействия использованию информационно-коммуникационных технологий в преступных целях;
- ✓ обеспечение безопасного и стабильного функционирования и развития информационно-телекоммуникационной сети «Интернет» на основе равноправного участия государств в управлении данной сетью и недопущению установления иностранного контроля над ее национальными сегментами;
- ✓ принятие политико-дипломатических и иных мер, направленных на противодействие политике недружественных государств по милитаризации глобального информационного пространства, по использованию информационно-коммуникационных технологий для вмешательства во внутренние дела государств и в военных целях, а также по ограничению доступа других государств к передовым информационно-коммуникационным технологиям и усилению их технологической зависимости.

Такой подход отличается от стратегий западных стран, ограничивающихся рамками «кибербезопасности», представляющей собой свод процессов, передовых практик и технологий, которые помогают защитить критически важные системы и сети от цифровых атак.

Россия выступает за широкий подход к содержанию данного понятия, включая в него как технические аспекты (безопасность цифровых технологий, информационных систем и сетей), так и значительный круг политико-идеологических вопросов (манипулирование информацией, пропаганда в глобальных информационных сетях, информационное воздействие). Страны коллективного Запада и США придерживаются узкого подхода, ограничиваясь технической стороной, используя термин «кибербезопасность». Учитывая специфику современных информационных отношений, многоаспектность рисков и угроз в информационном пространстве, позицию России следует признать более обоснованной и нацеленной на системный подход в вопросах регулирования обеспечения информационной безопасности.

В условиях цифровизации возникают новые информационные вызовы и угрозы, многие из которых носят трансграничный (глобальный) характер [11]. Как отмечается в докладе экспертов МГИМО, предложенный и продвигаемый Россией термин «международная информационная безопасность» подразумевает наличие не только технических, но и политико-идеологических угроз в данной области, что не коррелирует с западной концепцией кибербезопасности, акцентирующей внимание на технологическом измерении информационных угроз [12].

В России принято несколько правовых документов в сфере регулирования искусственного интеллекта,

25 Capabilities and risks from frontier AI, AI Safety Summit, October 2023. URL: https://assets.publishing.service.gov.uk/media/65395abae6c968000daa9b25/ frontier-ai-capabilities-risks-report.pdf.

в которых вопросам безопасности ИИ отводится важное место. Так, в 2019 году утверждена специальная Национальная стратегия развития искусственного интеллекта на период до 2030 года[26], которая провозглашает целями развития искусственного интеллекта 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 искусственного интеллекта.

Литература

1. *Ghazinoory, S., Fatemi, M., Saghafi, F. et al. A Framework for Future-Oriented Assessment of Converging Technologies at National Level. Nanoethics 17, 8 (2023). https://doi.org/10.1007/s11569-023-00435-4.*
2. *Мохов А. А. Демографическая безопасность и ее правовое обеспечение // Юрист. 2023. N6. С. 62–67.*
3. *Amatova, N. E.: Social consequences of the implementation of NBIC-technologies: risks and expectations. Univ. Soc. Sci. 9(8) (2014). 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, N10. 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 // Colliers 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 интелекта // 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.*

26 Утв. Указ Президента РФ от 10.10.2019 N 490 «О развитии искусственного интеллекта в Российской Федерации» // Собрание законодательства РФ, 14.10.2019, N 41, ст. 5700