УДК 34.09; 341 DOI: 10.31249/rgpravo/2026.01.07

СКУРКО Е.В.[1] **БЕЗОПАСНОСТЬ, КИБЕРБЕЗОПАСНОСТЬ ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ПРАВОВЫЕ АСПЕКТЫ (Обзор)**

Аннотация. Искусственный интеллект повсеместно в мире приобретает все более важную роль в общественной жизни. Системы ИИ, несмотря на их безусловные возможности, подвержены различным уязвимостям, которые могут поставить под угрозу безопасность их применения. Эти уязвимости, помимо этики ИИ, можно разделить на два основных типа: случайные сбои и преднамеренные атаки. Понимание этих потенциальных уязвимостей имеет основополагающее значение для разработки надежных и устойчивых механизмов противодействия, в особенности в правовой сфере. В обзоре анализируется безопасность и кибербезопасность применения ИИ в различных аспектах, юридические принципы и подходы к ответам на современные вызовы ИИ, рассматриваются международные акты и примеры актов национального законодательства, регулирующих использование ИИ в различных сферах социальной жизни и отношений.

Ключевые слова: искусственный интеллект; правовое регулирование; безопасность искусственного интеллекта; кибербезопасность.

SKURKO E.V. Security, cybersecurity of artificial intelligence applications: legal aspects (Review)

Abstract. Artificial intelligence becomes increasingly important in public life all-over the world. AI systems, despite their unconditional capabilities, are subject to various vulnerabilities that may compromise the security of their use. In addition to the ethics of AI, these vulner-

[1] *Скурко Елена Вячеславовна*, старший научный сотрудник отдела правоведения ИНИОН РАН, кандидат юридических наук.

abilities can be divided into two main types: accidental failures and deliberate attacks. Understanding these potential vulnerabilities is fundamental to develop reliable and sustainable countermeasures, especially in the legal field. The review analyzes the security and cybersecurity of the use of AI in its various aspects, legal principles and approaches responding to contemporary AI challenges, and examines international acts and examples of national legislation regulating the use of AI in various areas of social life and relationships.

Keywords: artificial intelligence; legal regulation; artificial intelligence security; cybersecurity.

Для цитирования: Скурко Е.В. Безопасность, кибербезопасность применения искусственного интеллекта: правовые аспекты (Обзор) // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Сер. 4: Государство и право. – 2026. – № 1. – С. 109–124. – DOI: 10.31249/rgpravo/2026.01.07

Введение

Искусственный интеллект сегодня выполняет сложные функции в качестве основы для принятия решений человеком в различных сферах жизни, либо в сфере управления техническими процессами – без участия человека. Такого рода «автономные системы», как ИИ, создают проблемы для правовой системы. Они непригодны в качестве носителей юридических обязанностей и прав. В результате, заменяя людей, они оставляют пробелы в правовой сфере, особенно в области юридической ответственности. По мнению ряда специалистов, законодатели смогут устранить эти пробелы инновационным способом только в том случае, если регулирование искусственного интеллекта будет специфичным для каждой отдельной области его применения, как например это частично достигнуто в сфере автономного транспорта и вождения [2, p. 9].

Потенциальные уязвимости в системах искусственного интеллекта и кибербезопасность

Системы искусственного интеллекта, несмотря на их большие и многоплановые способности, подвержены различным рискам и уязвимостям, которые могут создать угрозу их безопасности. Авторы книги «Агентский искусственный интеллект: теории и практики» под редакцией Кен Хуанга [1] разделяют уязвимости

ИИ на два основных типа: случайные сбои и преднамеренные атаки. Понимание этих угроз позволяет разрабатывать формы противодействия им – как техническими, так и юридическими средствами и методами [1, p. 369].

Случайные сбои в системах ИИ могут возникать из-за множества факторов, часто из-за непреднамеренных ошибок в их разработке, реализации или эксплуатации. Это:

– программные и логические ошибки представляют собой один из источников уязвимостей в системах ИИ. По мере того, как технологии ИИ становятся все более сложными, совершенствуются алгоритмы принятия решений и расширяются возможности обучения и взаимодействия различных систем ИИ, вероятность ошибок при принятии решений резко возрастает. Эти проблемы могут приводить к неожиданным системным сбоям и к серьезным последствиям в критически важных областях;

– аппаратные сбои. «Агенты» ИИ, особенно те, которые встроены в физические системы, такие как мобильные телефоны, умные часы, другие носимые устройства, умный дом, человекоподобные роботы и т.п., используют различные аппаратные компоненты для распознавания данных, их обработки и приведения прибора в действие. Сбои в работе этих компонентов могут нарушить способность «агента» точно воспринимать окружающую среду или выполнять действия по назначению;

– проблемы с качеством данных и предвзятость. Модели машинного обучения, используемые ИИ, хороши ровно настолько, насколько точны данные, на которых они обучены, а недостатки в данных обучения могут привести к предвзятому или неверному принятию решений. Эта уязвимость особенно важна для «агентов» ИИ, поскольку они часто работают автономно в динамичных средах, где последствия предвзятых решений со временем могут усугубиться [1, p. 370–372].

Преднамеренные атаки. Помимо случайных сбоев, системы ИИ сталкиваются с угрозами преднамеренных атак, направленных на использование их уязвимостей или манипулирование их поведением. Это:

– перехватывающие атаки на модели ИИ представляют значительную угрозу для систем ИИ, поскольку они могут манипулировать восприятием «агента» и процессами принятия решений. Эти атаки включают в себя создание входных данных, специально разработанных для того, чтобы обмануть системы ИИ и заставить их принимать неверные решения;

– атаки с использованием вредоносных данных – еще одна форма преднамеренного манипулирования ИИ. В ходе этих атак злоумышленники вводят поврежденные или вводящие в заблуждение данные в набор обучающих данных системы ИИ или в потоки онлайн-обучения. Для «агентов» ИИ, которые постоянно обучаются и адаптируются к своей среде, заражение данных представляет собой серьезную долгосрочную угрозу. Злоумышленник может постепенно влиять на поведение агента, вводя вредоносные данные;

– хищение моделей и обратное проектирование представляют угрозу для интеллектуальной собственности и безопасности систем ИИ. Злоумышленники могут попытаться извлечь базовую модель или ее параметры с помощью различных методов, включая атаки с использованием инверсии модели или логического вывода о ее принадлежности. Для «агентов» ИИ риски хищения моделей выходят за рамки проблем, связанных с интеллектуальной собственностью, поскольку извлеченные модели могут быть использованы для разработки более эффективных перехватывающих атак на систему ИИ; похищенные модели могут быть использованы, чтобы прогнозировать стратегии агента и противодействовать им; для «агентов» ИИ, работающих с конфиденциальными данными (например, в сфере здравоохранения или финансов), кража модели потенциально может привести к раскрытию персональных данных и т.п.;

– атаки на визуальные и языковые модели посредством всплывающих окон. Автономные системы ИИ, работающие на базе больших визуальных и языковых моделей (Vision and Language Models, VLM), показали значительные перспективы в выполнении различных задач, включая просмотр веб-страниц для бронирования поездок или управление программным обеспечением в персональном компьютере. Эти задачи требуют от систем ИИ понимания графических пользовательских интерфейсов и эффективного взаимодействия с ними, что обеспечивается интеграцией визуальной и лингвистической обработки. Поскольку визуальные входные данные становятся все более важными для приложений с ИИ, понимание рисков и уязвимостей, связанных с такими системами, становится критическим. При этом последствия такой визуальной интеграции для безопасности остаются недостаточно изученными [1, p. 373–375].

Безопасность ИИ в таких случаях – часть предмета кибербезопасности. Кибербезопасность организуется технически, а также все больше гарантируется юридически. Так, в современном взаи-

мосвязанном мире международные подходы и национальное нормативное регулирование в сфере кибербезопасности выступает на передний план. Как указывают эксперты в сфере кибербезопасности Джейсон Эдвардс и Гриффин Вивер, в значительной степени это связано с тем, что компании и организации, в том числе коммерческой направленности, по мере своего развития расширяют свое цифровое присутствие за пределами национальных границ, работая одновременно в нескольких юрисдикциях. Это расширение подпитывается многочисленными достижениями в области цифровых технологий, которые делают трансграничные операции не просто стратегическим преимуществом, но и необходимостью в мире, где конкуренция и сотрудничество больше не ограничены географическими рамками [3, p. 299].

Нормативные правовые акты в области кибербезопасности приняты и действуют во многих странах, и каждый из них разработан в соответствии с потребностями, вызовами и культурными особенностями соответствующих стран. Законы и другие нормативные правовые акты применяются правительствами государств для обеспечения безопасности цифровых операций в пределах своей юрисдикции и защиты персональных данных своих граждан. Эти законы могут охватывать широкий спектр таких вопросов, как конфиденциальность персональных данных, суверенитет данных, уведомление об их утечке и требования к мерам кибербезопасности и др. Главной целью является создание более надежной цифровой среды, в которой предприятия могут безопасно работать, а граждане – доверять цифровой экономике [3, p. 300].

При активном участии Российской Федерации была разработана и принята резолюцией 79/243 ГА ООН от 24.12.2024 г. Конвенция ООН против киберпреступности; укрепление международного сотрудничества в борьбе с определенными преступлениями, совершаемыми с использованием информационно-коммуникационных систем, и в обмене доказательствами в электронной форме, относящимися к серьезным преступлениям (далее – Конвенция). Хотя Конвенция не содержит прямых положений, регулирующих вопросы безопасности ИИ, она создает основу для решения ряда важных аспектов эффективного регулирования безопасного ИИ. Целями данной Конвенции являются: «a) содействие принятию и укрепление мер, направленных на повышение эффективности и результативности предупреждения киберпреступности и борьбы с ней; b) поощрение, облегчение и укрепление международного сотрудничества в предупреждении киберпреступности и борьбе с

ней; и c) поощрение, облегчение и поддержка технической помощи и создания потенциала в целях предупреждения киберпреступности и борьбы с ней, особенно в интересах развивающихся стран» (ст. 1).

Конвенция, «если иное не указано в ней, применяется: a) к предупреждению и расследованию уголовных правонарушений, признанных таковыми в соответствии с настоящей Конвенцией, и преследованию за них, включая замораживание, арест, конфискацию и возвращение доходов от таких правонарушений; b) к сбору, получению, сохранению и передаче доказательств в электронной форме для целей уголовного расследования или судопроизводства, как это предусмотрено в статьях 23 и 35 настоящей Конвенции» (ст. 3).

Пока рано делать прогнозы и выводы об эффективности предложенных Конвенцией подходов и методов, однако ее потенциал в сфере борьбы с киберпреступностью в мире несомненен.

Безопасное применение искусственного интеллекта в правовой сфере

Стефан Майер из Технического университета прикладных наук Вилдау (Германия) (The Technical University of Applied Sciences) рассматривает «правовые вызовы» ИИ на примере трех сфер его применения в правовой системе: в сфере государственного управления и судебной деятельности; в сфере автономного транспорта; в сфере интеллектуальных прав [4, p. 9–22].

1. *Искусственный интеллект в сфере государственного управления и в судебной деятельности.* Уровень развития и потенциал ИИ делают его привлекательным для использования в государственном управлении, а также в судебной деятельности, правоохранительными органами, юридическим сообществом. Так, ИИ более десятилетия активно привлекается в качестве системы поддержки принятия решений правоохранительными органами в разных странах. Например, в ФРГ с 2019 г. судебные органы земли Северный Рейн-Вестфалия реализуют проект по проверке изымаемых файлов на содержание детской порнографии с помощью ИИ-технологий. Тем не менее автоматически отобранные файлы, которые, по оценке ИИ, могут содержать информацию об уголовном преступлении, должны подвергаться окончательной проверке человеком. Как отмечает автор, перспективы использования ИИ в правоохранительной деятельности – это достижение момента, ко-

гда оценки ИИ будут составлять окончательное решение, т.е. ИИ будет окончательно определять доказанность по составу преступления и определять вытекающие из этого правовые последствия [4, p. 20].

Огромные исследовательские усилия сегодня направлены на то, чтобы алгоритмически «имитировать» понимание текста и его юридический анализ в целях решения задач, стоящих перед органами судебной власти и практикующими юристами [ibid.].

Учитывая риски в связи с применением ИИ в правовой сфере, в законодательстве ряда государств предусматривается прямой запрет безусловного применения ИИ в правовой практике, например ст. 35a Закона об административном судопроизводстве Германии (Verwaltungsverfahrensgesetz (BVwVfG) = Administrative Procedures Act) [Ibid]. Аналогичным образом, например, ст. 22 Общего регламента ЕС по защите данных (General Data Protection Regulation) (GDPR) и ст. 11 Директивы (ЕС) 2016/680 о защите физических лиц в связи с обработкой персональных данных компетентными органами в целях предотвращения, расследования, выявления или судебного преследования уголовных преступлений или исполнения уголовных наказаний (Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties) по общему правилу запрещают «автоматизированное индивидуальное принятие решений».

По мнению С. Майера, такой консервативный подход со временем станет нежизнеспособным. Учитывая потенциал ИИ, процесс принятия решений с помощью ИИ в будущем позволит избежать предвзятости со стороны человека, а также обеспечить единообразное применение закона при принятии юридически значимых решений [4, p. 21].

В связи с расширяющимся применением ИИ в правовой сфере экспертов и общественность беспокоят: во-первых, непредсказуемость функционирования ИИ; во-вторых, непрозрачность решений ИИ – как факторы, ограничивающие его широкое использование.

Вместе с тем, по утверждению С. Майера, чем больше ИИ будет «применять» закон, тем в большей степени он окажется способен соблюдать конкретные юридические требования при осуществлении властных полномочий, и, хотя «непредсказуемость» ИИ

может поставить под угрозу законность решений, которые он потенциально способен сформулировать, чиновники тоже подвержены риску принятия незаконных решений – будь то на основе ошибки либо в силу личной мотивации, которых нельзя исключить даже при самой тщательной подготовке государственных служащих. То есть, ограничение использования ИИ для осуществления государственной власти в силу «непредсказуемости» этой технологии может быть обосновано только тщательным сравнительным анализом склонности человека, с одной стороны, и ИИ – с другой, к ошибкам в правоприменении [ibid.].

Фактор «непрозрачности» ИИ составляет основание для ограничения использования ИИ в правовой сфере, поскольку, по общему правилу, правореализующие решения должны быть юридически понятными и поддающимися проверке. Поэтому, например, государственные органы должны дополнять принимаемые ими решения изложением причин, их обосновывающих. Однако алгоритмический путь, который выбирает самообучающийся ИИ для достижения искомого результата, технически невозможно сделать прозрачным во всей полноте. По этой причине «алгоритмическое управление» критикуется юристами, а в области компьютерных наук предпринимаются попытки решить эту проблему «непрозрачности» ИИ путем проведения исследований в области «объяснимого ИИ» («Explainable AI»)[1] [4, p. 21].

Вывод, к которому приходит С. Майер, – препятствия к безопасному применению ИИ в правовой сфере в принципе преодолимы, если не относиться к нему как «постоянному препятствию» для использования ИИ органами власти различного уровня и компетенции [Ibid].

2. *Применение искусственного интеллекта в автономном вождении.* В качестве интересного примера вызовов, возникающих вследствие применения ИИ, С. Майер выделяет проблематику автономного вождения в сфере дорожного движения на дорогах общего пользования. Это связано с заинтересованностью в широком внедрении автономного транспорта в повседневную жизнь, с одной стороны, и дискуссионностью вопроса юридической ответ-

[1] *Объяснимый искусственный интеллект* – набор процессов и методов, которые позволяют пользователям-людям понимать и доверять результатам и выводам, созданным алгоритмами машинного обучения (см.: Что такое объяснимый ИИ? – URL: https://www.ibm.com/think/topics/explainable-ai) (дата обращения: 15.11.2025).

ственности в случае ДТП с участием автономных транспортных средств – с другой.

Вопросы массового внедрения автономных ТС в дорожном движении, а также юридической ответственности в случае ДТП с их участием находятся в фокусе внимания законодательных органов многих стран мира. Тем не менее в большинстве юрисдикций, если автономное ТС становится причиной аварии с травмами и / или материальным ущербом из-за нарушения правил дорожного движения, уголовная ответственность не предусмотрена: пользователь системы автономного ТС ей не управляет, а, следовательно, не несет ответственности за небрежность; производитель, по крайней мере в том случае, если возник «просто» риск автономного вождения, не несет уголовной ответственности, поскольку такие риски считаются «допустимыми». Аналогичным образом не существует третьих сторон, которые могли бы быть привлечены к уголовной ответственности за нарушение правил дорожного движения [4, p. 19]. С. Майер проводит – в этом случае – аналогию с естественным причинам повреждения (например, смерть от удара молнии на открытом воздухе). Так, в случае естественных причин общество не ожидает, что кто-то будет привлечен к уголовной ответственности. В отличие от этого, подчеркивает автор, автомобиль – это артефакт, который используется в интересах отдельных людей. Поэтому сомнения в приравнивании его к «естественным» причинам повреждений представляются обоснованными. Возникает вопрос, готово ли общество признать, что никто не должен нести уголовной ответственности в случае серьезных дорожно-транспортных происшествий, вызванных автономным транспортным средством?

С. Майер полагает, что это – очевидный пробел в законодательстве об уголовной ответственности, и он может стать косвенным препятствием для развития и внедрения инноваций вообще, не говоря о том, что является существенным упущением законодателя и требует скорейшего устранения в глазах общественного мнения и этики [ibid.].

Решением отчасти может быть обязанность производителя постоянно контролировать свои автономные системы вождения после их появления на рынке (и дорогах). Нарушение этой обязанности, по мнению С. Майера, может получить квалификацию уголовного преступления. Кроме того это послужит дополнительной мотивацией для производителя надлежащим образом проводить мониторинг, что, в свою очередь, дополнительно повысит вероят-

ность устранения выявленных недостатков конструкции (например, путем обновления программного обеспечения и т.п.). В итоге предлагаемый подход к назначению уголовной ответственности, как утверждает С. Майер, дополнительно будет скорее способствовать инновациям, чем препятствовать им [ibid.].

3. *Искусственный интеллект и интеллектуальные права.* По общему правилу, законодательство большинства стран мира требует факта личного интеллектуального творчества (человека) – для того чтобы какое-либо произведение получило защиту авторского права. Например, Европейский суд (European Court of Justice) рассматривает это требование как принцип законодательства ЕС.

Согласно законодательству Германии, для получения правовой защиты произведения должны быть «личными интеллектуальными творениями», о чем говорит раздел 2(2) Закона об авторском праве и смежных правах 1965 г. (Urheberrechtsgesetz (UrhG) = Act on Copyright and Related Rights, UrhG). То есть, для признания авторского права требуется творческая активность конкретного человека, интеллектуальный и личный труд автора (соавторов). При этом, даже если отдельные компоненты произведения были созданы автоматически, оно может заслуживать правовой охраны.

В последнее время, однако, как обращает внимание С. Майер, появляется все больше «художественных» и «литературных» произведений, созданных исключительно ИИ. Например, «Следующий Рембрандт»[1] («The Next Rembrandt») – произведения, созданные самообучающимся ИИ на основе подражания стилю этого художника. В качестве обучающих данных для них использовалась коллекция картин Рембрандта. Еще один пример: в 2018 г. сгенерированная с помощью ИИ картина «Эдмонд де Белами» была продана на аукционе за 432 тыс. долл.[2] Она была сформирована из анализа около 15 тыс. картин, созданных в период с XIV по XIX в.

Оценка охраноспособности таких «произведений» ИИ, как указывает С. Майер, на практике схожа с решениями, принимающимися в отношении произведений, созданных случайным генера-

[1] The Next Rembrandt. – URL: https://www.vml.com/work/next-rembrandt (дата обращения: 15.11.2025).

[2] Obvious and the Interface Between Art and Artificial Intelligence: As Christie's Becomes the First Auction House to Offer an Artwork Created by an Algorithm, We Ask if AI is Set to Become Art's Next MEdium. 12 December 2018. – URL: https://www.christies.com/en/stories/a-collaboration-between-two-artists-one-human-one-a-machine-0cd01f4e232f4279a525a446d60d4cd1 (дата обращения: 15.11.2025).

тором: согласно сложившимся подходам, такого рода произведения, чтобы получить правовую охрану, требуют определенного уровня человеческого участия, т.е. интеллектуального вклада человека в конечный продукт. Например, автор может создать несколько шаблонов, которые затем обрабатываются при помощи генератора случайных чисел. Работа самого генератора случайных чисел при этом расценивается лишь как имитация шаблонов художника-автора, т.е. стиля данного художника [4, p. 11].

Сам по себе стиль как таковой не подлежит правовой охране, и конкретное творческое влияние человека на конечный продукт, как например в вышеописанном случае, отсутствует. На практике зачастую применяется метод введения дополнительного требования к произведению, чтобы признать его соответствующим уровню охраноспособности: «автор» должен выбрать, какой из различных результатов работы генератора случайных чисел может быть представлен в качестве «произведения искусства». Другой подход состоит в том, что достаточно простого предъявления «произведения» – «объекта», который кажется «художественным», – чтобы предоставить по нему защиту авторских прав. С. Майер указывает на еще один существующий подход, при котором защита авторских прав на «произведение» ИИ будет предоставляться, если данные для обучения ИИ при создании этого произведения были основаны на работах художника – человека, который использует данное приложения ИИ, и т.п. [ibid.].

Приведенные выше примеры «произведений» ИИ, – «Следующий Рембрандт» и «Эдмонд де Белами», – как указывает С. Майер, не отвечают ни одному из приведенных подходов, предоставляющих защиту авторских прав произведениям, выполненных средствами ИИ. Эффективное правовое признание «произведений» ИИ охраноспособными, по мнению С. Майера, – принятие в юридической практике так называемой «доктрины отбора» (т.е. когда человек выбирает, какое из сгенерированных ИИ «произведений» заслуживает таковым являться). Сегодня, однако, приходится признать, что защита авторских прав в отношении «произведений», созданных с помощью ИИ, практически отсутствует [Ibid].

Актуальный вопрос для правоведов состоит в том, должны ли в принципе «произведения», созданные с помощью ИИ, подпадать каким-либо образом под авторское право: очевидно, что «творческий ИИ» будет становиться все более мощным и, возможно, в ближайшее время выйдет за рамки простой эклектики. С. Майер подчеркивает, что законодатель сегодня располагает

значительной свободой действий в сфере распределения интеллектуальных прав по различным аспектам инновационной политики, включая произведения ИИ [4, p. 12].

Правовые решения по безопасности искусственного интеллекта

С. Майер полагает, что имеющиеся проблемы и недостатки как в части правотворчества, так и в плане правоприменения в отношении ИИ и вопросах его безопасного использования, подразумевают, что в развитии законодательства об ИИ требуется четко учитывать два аспекта: 1) соразмерность интересов третьих лиц в ИИ и гарантий базовых прав граждан; 2) оценка рисков от внедрения технологий ИИ [4, p. 22–23].

Так, на первый взгляд, развитие правового регулирования в сфере технологий предполагает обеспечение баланса основных экономических прав производителей и распространителей технологий в отношении основных прав тех, кому такие технологии могут причинить вред, будь то права на защиту персональных данных или физическую неприкосновенность и др.

Однако новые технологии выгодны не только производителям и дистрибьюторам, но и массам их пользователей – в повседневной жизни в том числе. Этот аргумент, безусловно, признается регулирующими органами и порой используется в качестве обоснования в поддержку внедрения новых технологий (например, снижение смертности на дорогах благодаря автономному вождению). При этом законодатели уделяют недостаточно внимания ограничивающему регулированию технологий, когда юридически оказывается допустимым, что будут ущемлены основные права (например пациенты, которые умирают из-за того, что медицинский ИИ, который мог бы в противном случае спасти им жизнь, стал доступен слишком поздно). Неопределенность относительно того, действительно ли новая технология могла быть доступна ранее в отсутствие ее нормативного регулирования, не может служить аргументом против юридической значимости причинения вреда третьей стороне, вызванного этим (отсутствием) регулированием, полагает С. Майер. Речь идет о принципе предосторожности, основная цель которого, особенно в вопросе о применении технологий ИИ, состоит в том чтобы разрешать применение этой технологии до того, как будет выявлен и установлен (потенциальный) вред, ею наносимый [4, p. 23].

Однако, в том числе на международном уровне, получил распространение «подход, основанный на оценке рисков» (risk-based approach), которому следует, например, Евросоюз. Этот подход имеет два элемента. Во-первых, требуется научное подтверждение потенциальной способности технологии причинять вред («опасность»). Во-вторых, должна существовать некоторая вероятность того, что юридический актив, который должен быть защищен нормативным актом, действительно может подвергнуться такой опасности в связи с планируемым использованием технологии («воздействие»). Поэтому требуется оценка конкретного использования технологии («оценка воздействия»).

Следуя этой «схеме», например, Евросоюз, как правило, ориентируется на создание процедуры допуска продукта на рынок, а не на свойства и качество самого продукта. Это относится и к системам и технологиями ИИ, которые, если будут отнесены к категории «высокого риска», будут регулироваться наложением множества стандартных нормативных обязательств на поставщиков, импортеров, дистрибьюторов и пользователей [4, p. 24].

Безопасность и конфиденциальность искусственного интеллекта в стандартах и правилах международных организаций

Авторы книги «Понимание принципов работы ИИ в сфере кибербезопасности и безопасного ИИ: проблемы, стратегии и тенденции (прогресс в сфере ИИ)» [5] – Дилли Прасад Шарма из Университета Торонто (University of Toronto), Араш Хабиби Лашкари, Йоркский университет (York University), Махди Дагмехчи Фирузджаи, Университет Макьюэн (MacEwan University), Самане Махдавифар, Университет Макгилла (McGill University), Пулей Сюн, Национальный исследовательский совет Канады (National Research Council of Canada) – в своем исследовании обращают внимание на то, что безопасность и конфиденциальность ИИ стали важнейшими приоритетами в развитии нормативного регулирования, по мере того как технологии ИИ все больше интегрируются в глобальные отрасли и повседневную жизнь. В связи с этим были разработаны определенные международные стандарты и нормативы, дающие рекомендации по снижению рисков и обеспечению этичного использования ИИ – для государств и частных компаний и организаций [5, p. 216–220]. Это, в частности, проекты ISO – ISO/IEC DIS 27090 «Кибербезопасность – Искусственный интеллект – Руководство по устранению угроз безопасности и компро-

метации систем искусственного интеллекта»[1]; ISO/IEC DIS 27091 «Кибербезопасность и конфиденциальность – Искусственный интеллект – Защита конфиденциальности»[2]. Эти стандарты охватывают все этапы жизненного цикла ИИ, от сбора данных до системной интеграции, обеспечивая надежное руководство для государств и организаций, соответствующее глобальным требованиям безопасности и этики.

Важную роль в разработке стандартов, обеспечивающих безопасное и прозрачное внедрение технологий ИИ, играет Европейский институт по стандартизации в области телекоммуникаций (European Telecommunications Standards Institute) (далее – ETSI). Рекомендации ETSI особенно востребованы для таких отраслей, как телекоммуникации, где надежность систем ИИ имеет первостепенное значение. Они направлены на предотвращение злоупотреблений и уязвимостей, повышают доверие к приложениям ИИ, используемым для критически важной инфраструктуры и сервисов.

Технический комитет ETSI по защите искусственного интеллекта (TC SAI) ориентируется в своей работе на то, чтобы повысить безопасность ИИ путем разработки высококачественных технических стандартов и рассматривает четыре основных аспекта стандартизации безопасности ИИ: 1) техническая защита ИИ от атак; 2) смягчение последствий, вызванных техническими проблемами с ИИ; 3) применение ИИ для противодействия техническим атакам; 4) аспекты общественной безопасности при использовании и применении ИИ[3].

Работы по стандартизации ETSI адресованы всем заинтересованным сторонам, включают конечных пользователей, производителей, операторов и правительства [5, p. 214].

Принципы ИИ ОЭСР (OECD AI Principles[4]) представляют собой первый межправительственный стандарт в области ИИ, ус-

[1] ISO/IEC DIS 27090 Cybersecurity – Artificial Intelligence – Guidance for Addressing Security Threats and Compromises to Artificial Intelligence Systems. – URL: https://www.iso.org/standard/56581.html (дата обращения: 15.11.2025).

[2] ISO/IEC DIS 27091 Cybersecurity and Privacy – Artificial Intelligence – Privacy protection. – URL: https://www.iso.org/standard/56582.html (дата обращения: 15.11.2025).

[3] Securing Artificial Intelligence (SAI). – URL: https://www.etsi.org/technologies/securing-artificial-intelligence (дата обращения: 15.11.2025).

[4] OECD AI Principles. – URL: https://www.oecd.org/en/topics/ai-principles.html (дата обращения: 15.11.2025).

танавливающий глобальный ориентир для разработки и внедрения надежных систем ИИ.

Принципы ИИ ОЭСР основаны на пяти ключевых ценностях: 1) инклюзивный рост, устойчивое развитие и благополучие: ИИ должен вносить позитивный вклад в прогресс общества, поддерживая устойчивое развитие и повышая благосостояние; 2) ориентированные на человека ценности и справедливость: системы искусственного интеллекта должны уважать человеческое достоинство и права личности, обеспечивая справедливость при их разработке и применении; 3) прозрачность и объяснимость: системы искусственного интеллекта должны работать прозрачно и их решения должны быть понятны, укреплять доверие и подотчетность; 4) надежность и защищенность: системы искусственного интеллекта должны быть устойчивыми, защищенными и безотказными, сводя к минимуму риски и обеспечивая безопасную эксплуатацию; 5) подотчетность: организации и разработчики должны нести ответственность за воздействие и конечные результаты своих систем искусственного интеллекта, обеспечивая эффективный надзор [5, p. 216–217].

Чтобы содействовать реализации этих принципов, ОЭСР предлагает пять практических рекомендаций для правовой политики в сфере ИИ национальных государств: 1) инвестиции в исследования и разработки в области искусственного интеллекта; 2) содействие созданию цифровой экосистемы для ИИ; 3) формирование благоприятной политической среды; 4) наращивание человеческого потенциала и подготовка к трансформации рынка труда; 5) международное сотрудничество для обеспечения надежности ИИ [5, p. 217].

Заключение

Безопасность применения ИИ в общественной и правовой жизни определяется, с одной стороны, особенностями систем ИИ, с другой – особенностями сфер их приложения в социальной практике.

Системы ИИ могут создавать социальные риски, но и сами при этом подвержены рискам и уязвимостям, что ставит под угрозу их безопасность. Уязвимости ИИ в техническом плане подразделяются на: случайные сбои и преднамеренные атаки. Риски ИИ в социально-правовой сфере определяются сферой применения ИИ.

Существует экспертное мнение, что в том случае, если регулирование искусственного интеллекта будет специфичным для каждой отдельной области его применения, как например это происходит в сфере автономного транспорта, угрозы и риски безопасности его эксплуатации могут быть существенно снижены, – с чем, в целом, можно согласиться.

Сегодня разрабатываются первые нормативные акты и международные стандарты в области безопасности применения ИИ, однако остается много вопросов в развитии правового регулирования применения ИИ как на национальном, так и на международном уровнях.

Список литературы

1. Agentic AI: Theories and Practices / ed. Ken Huang. – Cham: Springer, 2025. – 438 p.
2. Artificial intelligence in application: Legal aspects, application potentials and use scenarios / ed. T. Barton, C. Müller. – Wiesbaden: Springer, 2024. – 208 p.
3. Edwards J., Weaver G. The Cybersecurity Guide to Governance, Risk, and Compliance. – Hoboken: Wiley, 2024. – 672 p.
4. Meyer S. Legal Challenges of Artificial Intelligence and How to Manage Them // Artificial intelligence in application: Legal aspects, application potentials and use scenarios / ed. T. Barton, C. Müller. – Wiesbaden: Springer, 2024. – P. 9–30.
5. Understanding AI in Cybersecurity and Secure AI: Challenges, Strategies and Trends / D.P. Sharma, A.H. Lashkari, M.D. Firoozjaei, S. Mahdavifar, Pulei. Xiong. – Cham: Springer, 2025. – 255 p.