УДК 34.09 DOI: 10.31249/iajpravo/2026.01.08

АЛФЕРОВ О.Л.[1], АЛФЕРОВА Е.В.[2] ИНТЕГРАЦИЯ НАДЕЖНЫХ МЕХАНИЗМОВ ПРАВОВОЙ ЗАЩИТЫ В СИСТЕМУ УПРАВЛЕНИЯ ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ (Обзор)

Аннотация. В обзоре представлены точки зрения некоторых исследователей на взаимосвязи между нормативно-правовой базой, этическими концептами и развивающимися технологиями, на пробелы и противоречия в законодательстве в области управления ИИ и на механизмы защиты прав человека, предотвращения или смягчения потенциального вреда, связанного с ИИ. Анализ законодательства позволяет авторам предложить эффективные правовые меры против необоснованного вторжения в частную жизнь, дискриминационной практики, несправедливых решений в сфере жилья, уголовного правосудия и многих других областей. Предлагается четко урегулировать ответственность за вред, причиненной искусственным интеллектом, упорядочить правила возмещения вреда и устранения негативных последствий функционирования систем искусственного интеллекта.

Ключевые слова: искусственный интеллект; генеративный искусственный интеллект; правовая защита; юридическая ответственность; справедливость; закон об искусственном интеллекте; фундаментальные права; управление искусственным интеллектом; правовое регулирование технологий искусственного интеллекта.

ALFEROV O.L., ALFEROVA E.V. Integration of reliable legal protection mechanisms into the artificial intelligence management system (Review)

[1] *Алферов Олег Леонидович*, ведущий редактор отдела правоведения ИНИОН РАН.

[2] *Алферова Елена Васильевна*, ведущий научный сотрудник отдела правоведения ИНИОН РАН, кандидат юридических наук.

Abstract. The review presents the views of some researchers on the relationship between the regulatory framework, ethical concepts and emerging technologies, gaps and contradictions in AI management legislation, and mechanisms to prevent or mitigate potential harm associated with AI. The analysis of legislation allows the authors to propose effective legal measures of liability for damage caused by artificial intelligence, and compensation for damage, as well as elimination of negative consequences of the functioning of artificial intelligence systems.

Keywords: artificial intelligence; generative artificial intelligence; legal protection; legal responsibility; justice; law on artificial intelligence; fundamental rights; artificial intelligence management; legal regulation of artificial intelligence technologies.

Для цитирования: Алферов О.Л., Алферова Е.В. Интеграция надежных механизмов правовой защиты в систему управления искусственным интеллектом (Обзор) // Социальные и гуманитарные науки. Отечественная и зарубежная литература: ИАЖ. Сер. 4: Государство и право. – 2026. – № 1. – С. 125–141. – DOI: 10.31249/iajpravo/2026.01.08

Введение

Системы ИИ все больше интегрируются во многие ключевые отрасли, такие как здравоохранение, финансы, трудоустройство, уголовное судопроизводство, образование и другие. Преимущества, связанные с повышением эффективности их функционирования в эпоху ИИ, идут рука об руку со значительными проблемами, обусловленными многочисленными рисками нарушения основных прав и причинения вреда в результате необоснованного вторжения в частную жизнь, дискриминационной практики, несправедливых решений в сфере жилья, уголовного правосудия и многих других областей. К сожалению, люди, которые становятся жертвами правонарушений с помощью искусственного интеллекта, могут оказаться бессильными без эффективных средств правовой защиты их прав.

В данном обзоре представлены точки зрения некоторых исследователей – участников Кембриджского форума «Искусственный интеллект: право и управление» [2], указывающих на настоятельную потребность интеграции надежных механизмов правовой защиты основных прав человека в систему управления ИИ, поскольку технологии ИИ все шире внедряются в жизнь государства

и общества и представляют определенную угрозу. Технологии ИИ могут ограничивать личную свободу, разрушительно влиять на индивидуальное самоопределение; они уязвимы для некоторых людей, неспособных понять и работать с ИИ, и др. Управление ИИ должно защищать права и интересы всех людей, на которых влияют системы ИИ, регулировать способы предотвращения вреда, причиняемого ИИ, и обеспечивать возмещение ущерба, когда вред уже нанесен.

Индивидуальные меры защиты в эпоху искусственного интеллекта: справедливые алгоритмы, справедливое регулирование, справедливые процедуры

Традиционная правовая доктрина требует, чтобы для предотвращения или смягчения рисков нарушения основных прав были созданы новые меры защиты, а существующие – усилены. Защита основных прав лежит в основе международного права и национальных нормативных правовых актов и академических идей, связанных с технологиями ИИ. *Люпчо Гродзановский* и *Джером Де Куман Анно*т из Льежского университета (Бельгия) считают, что в условиях внедрения ИИ в разные сферы жизни возникает необходимость в усилении индивидуальных мер защиты с помощью реализации субъективных прав человека, закрепленных в законе [3].

Понятие «индивидуальная защита» авторы рассматривают как *нормативную цель защиты* индивидуальных прав и *право требовать* определенного вида защиты или охраны. Предполагается, что «справедливый» закон защищает свободу и равенство личности. В социальных и гуманитарных науках «справедливость» всегда рассматривалась как основополагающее, но неуловимое понятие. По мере того как технологии ИИ демонстрировали свою способность причинять вред, ученые поднимают вопросы о справедливости, которые не могли должным образом охватить стандартные нормативные и научные теории. Реакция специалистов по этике заключалась в том, чтобы вернуться к классике и заложить основы «новой» этической системы, в рамках которой в конечном итоге можно было бы ввести в действие законодательство об ИИ. Так называемая «классика» – это три основных направления в этике: добродетельное, утилитаристское и деонтическое [3].

Учитывая роль ИИ, Л. Гродзановский и Дж. Де Куман утверждают, что «защита личности» как цель *справедливого* регули-

рования оправдана именно потому, что технологии ИИ могут подорвать ключевые требования справедливости (свободу и равенство личности), порождая новые формы неравенства и ограничивая способность людей свободно и осмысленно действовать [3].

О каких угрозах идет речь?

Во-первых, *об ограничении личной свободы.* По мнению авторов, угроза того, что ИИ будет принуждать людей принимать решения, которые они вероятно не приняли бы, если бы обладали полной (свободной от ИИ) проницательностью, стала чемто обыденным.

Во-вторых, о *разрушительном влиянии ИИ на индивидуальное самоопределение и непрозрачности ИИ* и отсутствии должного человеческого контроля и надзора.

В-третьих, об *уязвимости некоторых людей из-за их неспособности понять и работать с новыми технологиями*. Конечно, не все люди одинаково осведомлены об угрозах, связанных с технологиями ИИ. Некоторые группы характеризуются особенностями, которые повышают вероятность того, что они станут жертвами манипуляций со стороны ИИ. В своей книге Джанклаудио Мальджери[1], на которую ссылаются авторы, выделил четыре архетипические «неспособности», которые характеризуют или усиливают уязвимость людей: неспособность понять информацию об обработке данных; неспособность понять риски, их значение и последствия; неспособность дать действительное согласие и неспособность надлежащим образом реализовать права на защиту данных (цит. по: [3]).

Среди примеров несправедливости, которую могут вызывать технологии ИИ, наиболее заметным и актуальным примером признается дискриминация, которая приводит к исключению и маргинализации определенных групп[2]. Есть ряд секторов, таких как биометрия, критически важная инфраструктура, образование и профессиональная подготовка, трудоустройство и доступ к основным государственным услугам, общей чертой которых является вероятность дискриминации.

В-четвертых, об *угрозе справедливости, исходящей от ИИ.* Здесь Л. Гродзановский и Дж. де Куман усматривают два ключе-

[1] Malgieri G. Vulnerability and Data Protection Law. – Oxford: Oxford Univ. Press, 2023. – 308 p.

[2] Ethics guidelines for trustworthy AI. – URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения: 11.10.2025).

вых аспекта. Первый – *инструментальный*: справедливость выступает как средство достижения индивидуальной защиты. Ключевую роль здесь играют концепции, заложенные в законы об ИИ. В качестве положительного примера авторы приводят Закон ЕС об ИИ (AI Act), положения которого направлены на защиту разумных и добровольных действий людей, а также их равного доступа к различным правам и льготам и пользования ими (например, право на объяснение, льготы в так называемых секторах повышенного риска).

Второе измерение взаимосвязи гарантий и справедливости – *консеквенциалистское*: при эффективном применении материальные и процессуальные индивидуальные гарантии направлены на достижение справедливых результатов. Действительно, материальные права (защита данных, неприкосновенность частной жизни, недискриминация и т.д.) и процессуальные права (доступ к средствам правовой защиты и правосудию, эффективное судебное возмещение ущерба) обеспечивают гарантии и защитные механизмы, на которые люди могут полагаться, чтобы либо предотвратить несправедливый исход (например нарушение основного права), либо исправить его, если он уже произошел (например, возместить причиненный ущерб) [ibid.].

Право на возмещение ущерба является основополагающим правом человека, необходимым для устранения нарушений прав и свобод. Важность механизмов возмещения ущерба закреплена в международных документах по правам человека, таких как ст. 8 Всеобщей декларации прав человека, ст. 2 Международного пакта о гражданских и политических правах и др.

Расширение возможностей управления искусственным интеллектом с помощью механизмов возмещения ущерба

Управление ИИ, подчеркивают *Юлу Пи* (Центр междисциплинарных методологий Уорикского университета (Великобритания)) и *Мэдди Проктор* (Центр социальных исследований Гарвардского университета (США), должно защищать права и интересы всех людей, на которых влияют системы ИИ, предусматривая возможность предотвращения вреда, причиняемого ИИ, и возмещения ущерба, когда вред уже нанесен. Обеспечить условия для устранения как индивидуального, так и коллективного вреда, причиненного ИИ, по их мнению, можно двумя важнейшими способами: 1) создание надежных механизмов возмещения ущерба,

которые предоставляют отдельным лицам и сообществам официальные возможности для получения компенсации или принятия корректирующих мер; и 2) гарантия возмещения ущерба для всех, кто пострадал от систем ИИ [4].

Возмещение ущерба – комплекс мер, направленных на устранение вреда или негативных последствий, с которыми сталкиваются отдельные лица или сообщества в результате неправомерных действий. Цель возмещения ущерба – устранить или исправить нежелательную или несправедливую ситуацию. Пострадавшие лица могут добиваться принятия этих мер различными способами, включая судебные механизмы (национальные или региональные суды, международные правозащитные организации), государственные внесудебные механизмы (регулирующие органы, омбудсмен, органы по рассмотрению жалоб) и внутренние механизмы подачи жалоб в компаниях. Хотя возмещение ущерба может ассоциироваться у людей с денежной компенсацией, результат процесса возмещения ущерба может принимать различные формы, включая реституцию, компенсацию, реабилитацию, удовлетворение и гарантии неповторения.

В контексте несправедливости, связанной с искусственным интеллектом, Ю. Пи и М. Проктор (со ссылкой на других исследователей) выделяют два типа результатов возмещения ущерба: *восстановительный и карательный.* Восстановительное возмещение ущерба направлено на «возмещение материального возмещения ущерба стороне или жертве в результате неправомерного действия в результате нарушения ее прав, в то время как карательное возмещение ущерба предполагает наказание правонарушителя, часто в судебном порядке. Доступ к механизмам возмещения ущерба способствует всестороннему расследованию нарушений прав человека и причиненного вреда, позволяя надлежащим образом устранять ущерб, выплачивать компенсацию жертвам и привлекать к ответственности виновных» [ibid.].

Для управления ИИ с целью защиты отдельных лиц и общества от потенциального вреда, причиняемого ИИ, авторы предлагают использовать два основных подхода: *упреждающий и постфактумный*. Различие между этими механизмами регулирования ИИ состоит в предмете спора и времени его возникновения. Так, механизмы ex-ante (с лат. – до события) – это перспективные инструменты, которые вступают в силу до развертывания системы ИИ и начала ее воздействия на пользователей, в то время как механизмы ex-post применяются после развертывания системы и начала ее

работы. Акцент на важности механизма возмещения ущерба как постфактумной меры по устранению вреда, связанного с ИИ, по мнению авторов, не призван умалить легитимность и значимость превентивных мер. Скорее, речь идет о необходимости комплексного подхода к защите общества от потенциальных рисков и последствий, связанных с ИИ. Предварительные меры, такие как оценка рисков, этические рекомендации, например Организации экономического сотрудничества и развития[1]; Организация Объединенных Наций по вопросам образования и организации[2] и технические стандарты проектирования и разработки Международной организации по стандартизации[3], играют решающую роль в понимании, предотвращении и смягчении вреда, связанного с ИИ. Эти меры помогают обеспечить ответственное проектирование, разработку и внедрение систем ИИ с целью минимизации вероятности негативных последствий [4].

Признавая важный пробел в комплексном подходе к защите общества от потенциальных рисков и последствий, Ю. Пи и М. Проктор отмечают появление существенных законодательных новелл, например в последней версии Закона ЕС об ИИ[4] и Белой книге о регулировании ИИ в Великобритании[5]. Эти акты, по их мнению, значительно усилили внимание к механизмам защиты и возмещения ущерба в ответ на вред, наносимый ИИ. Изначально Закон ЕС об ИИ, предложенный Еврокомиссией в апреле 2021 г.

[1] Organisation for Economic Co-operation and Development // OECD AI principles. – 2019. – URL: https://oecd.ai/en/ai-principles (дата обращения: 11.10.2025).

[2] United Nations Educational, S., & Organization, C. Recommendation on the ethics of artificial intelligence. – 2021. – URL: https://www.unesco.org/en/artificial-intelligence/recommendation-ethics (дата обращения: 11.10.2025).

[3] International Organization for Standardization. ISO/IEC TR 24027:2021 information technology – Artificial. – 2021. – URL: https://www.iso.org/standard/81230.html (дата обращения: 11.10.2025).

[4] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). – URL: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng (дата обращения: 11.10.2025).

[5] UK's Department for Science, I. and Technology. Implementing the UK's AI regulatory principles: Initial guidance for regulators. – 2024. – URL: https://assets.publishing.service.gov.uk/media/65c0b6bd63a23d0013c821a0/implementing_the_uk_ai_regulatory_principles_guidance_for_regulators.pdf (дата обращения: 11.10.2025).

(вступил в силу 1 августа 2024 г.), подвергался критике за то, что в нем не были должным образом учтены проблемы, связанные с правами человека, из-за отсутствия надежного механизма подачи жалоб и возмещения ущерба. Однако последующие изменения значительно улучшили ситуацию[1]. В пересмотренном Законе теперь учитываются права отдельных лиц. В частности, для гарантии этих прав в Закон ЕС об ИИ были добавлены ст. 85, 86 и 99(10). Статья 85 позволяет отдельным лицам или группам лиц подавать жалобы в органы надзора за рынком, если их права, предусмотренные Законом, нарушаются системой ИИ. Статья 86 обеспечивает право на получение разъяснений по результатам работы систем ИИ с высоким уровнем риска, которые влияют на законные права, здоровье, безопасность, социально-экономический статус или другие основные права. Статья 99(10) предусматривает эффективные средства правовой защиты и надлежащую правовую процедуру в отношении действий органов по надзору за рынком. Этих прав не было в первоначальном проекте закона, что является важным шагом на пути к эффективному индивидуальному возмещению ущерба [4].

Аналогичным образом в Белой книге по регулированию ИИ, «пропорциональной и ориентированной на инновации нормативно-правовой базе», опубликованной правительством Великобритании 29 марта 2023 г., подчеркивается, что состязательность и возмещение ущерба являются важнейшими принципами[2].

Ю. Пи и М. Проктор выделяют несколько обязательных шагов достижения успехов в споре и возмещению вреда. *Первый – инициирование процесса,* который начинается с определения конкретных задействованных систем ИИ и понимания причиненного ими вреда. Однако, предупреждают они, этот шаг часто оказывается сложным из-за недостаточной прозрачности использования ИИ и запоздалого признания его негативных последствий. Отдельная проблема видится в том, что люди нередко не осознают, что они стали участниками процесса принятия решений на основе ИИ или столкнулись с контентом, созданным ИИ. Системы ИИ

[1] Engler A. Key enforcement issues of the AI act should lead EU trilogue debate. – URL: https://www.brookings.edu/articles/key-enforcement-issues-of-the-ai-act-should-lead-eu-trilogue-debate/ (дата обращения: 11.10.2025).

[2] UK Department for Science, I. and Technology 2023. A pro-innovation approach to AI regulation. – URL: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper (дата обращения: 11.10.2025).

часто внедряются без публичного уведомления, из-за чего людям сложно идентифицировать эти системы и понять их последствия. Такое неосознание особенно проблематично в случаях ошибочных решений или институциональных нарушений. В связи с этим, полагают авторы, необходимо обеспечить введение правила информирования пользователей и заинтересованных лиц о том, как они должны взаимодействовать с системой ИИ. Такое правило имеет решающее значение, поскольку без него люди могут не понять, когда система ИИ причиняет вред, и им будет сложно начать процесс возмещения ущерба [4].

Еще одной серьезной проблемой, препятствующей началу процесса возмещения ущерба, является запоздалое признание вреда, причиненного ИИ.

Второй шаг после выявления ущерба – *определение наиболее подходящих способов возмещения ущерба*. Это могут быть внутренние механизмы подачи жалоб в компании, государственные внесудебные механизмы, такие как службы омбудсменов, и судебные органы.

Рассматривая проблемы, с которыми сталкиваются люди, стремящиеся к возмещению ущерба, Ю. Пи и М. Проктор на реальных примерах демонстрируют различные препятствия, возникающие на каждом этапе возмещения ущерба в ЕС и США. Отмечается, что подходы к регулированию в ЕС и США часто сравнивают из-за различий в моделях управления ИИ: в ЕС особое внимание уделяется нормативному надзору, а в США предпочтение отдается более рыночному и децентрализованному подходу. Изучив эти разные системы, авторы замечают, что ни одна из них в полной мере не удовлетворяет потребность в эффективных средствах правовой защиты для лиц, пострадавших от систем ИИ. Несовершенство существующих механизмов правовой защиты подвергает людей значительным рискам, будь то утечка данных, аварии с участием автономных транспортных средств или алгоритмическая предвзятость.

Вывод исследователей: механизмы правовой защиты в условиях использования ИИ-технологий являются жизненно важными гарантиями, предоставляющими пострадавшим лицам возможность требовать компенсацию, исправления ошибок или, по крайней мере, пересмотр решений, принятых системами ИИ. Такие механизмы не только защищают права отдельных лиц, но и укрепляют доверие потребителей, что, в свою очередь, способст-

вует долгосрочному росту и этичному развитию индустрии ИИ [4].

Одним из таких механизмов является *омбудсмен по искусственному интеллекту.* Так, в Финляндии было принято примечательное решение омбудсмена, успешно восстановившее справедливость путем своевременного информирования потерпевшего лица и предотвращения дальнейшего нарушения его конфиденциальности. В 2020 г. Национальное бюро расследований Финляндии использовало программное обеспечение для распознавания лиц от Clearview AI для выявления потенциальных жертв сексуального насилия над детьми, не применяя надлежащих мер защиты конфиденциальности, таких как ограничения на срок хранения данных или передачу третьим лицам. Национальное управление полиции должно было уведомить о проекте финское Управление омбудсмена по защите данных, что они и сделали в 2021 г. после сообщения об утечке персональных данных. В ответ на это омбудсмен распорядился проинформировать пострадавших о взломе и удалить соответствующие персональные данные [4].

Важную роль в достижении справедливых результатов в случае крупномасштабного вреда, наносимого пользователям ИИ, играют регулирующие органы и службы защиты прав потребителей для получения компенсации. Например, Международная сеть по защите прав потребителей и правоприменению (International Network for Consumer Protection and Law Enforcement, ICPEN) – это международная организация, возглавляемая Федеральной торговой комиссией США (The Federal Trade Commission) (далее – FTC) и состоящая из 70 органов-членов. В США FTC обладает широкими полномочиями по возмещению ущерба, нанесенного ИИ. Раздел 5 Закона о FTC (Federal Trade Commission Act 1914), например, наделяет FTC полномочиями по регулированию недобросовестных и вводящих в заблуждение практик. Федеральная торговая комиссия может подавать в суд на компании и возмещать ущерб пострадавшим пропорционально, издавать судебные запреты, обязывающие компании прекратить вредоносную практику, или заключать долгосрочные соглашения о согласии, предусматривающие дальнейший мониторинг и штрафы. Так, в рамках проверки на предмет спам-звонков FTC в сотрудничестве с генеральными прокурорами штатов обратилась к поставщикам услуг в рамках «Операции по борьбе со спам-звонками» и объявила совместную операцию «Стоп спам-звонкам».

Во многих штатах США также действуют законы об ответственности за введение потребителей в заблуждение, например Закон о недобросовестной конкуренции в Калифорнии (the Unfair Competition Law, UCL), который позволяет отдельным потребителям подавать в суд с требованием о судебном запрете [ibid.].

Еще один шаг – *судебный пересмотр и возмещение ущерба в судебном порядк*е. Гражданское право предлагает несколько способов возмещения ущерба в судебном порядке, в том числе в связи с ответственностью за качество продукции и халатностью. Вместе с тем, как отмечают Ю. Пи и М. Проктор, в США до сих пор не было ни одного успешного иска о возмещении ущерба в связи с качеством программного обеспечения. 23 октября 2024 г. Европарламент и Совет утвердили новую Директиву (ЕС) 2024/2853 об ответственности за дефектную продукцию (Directive (EU) on liability for defective products), которая распространяет ответственность за дефектную продукцию на цифровые продукты и программное обеспечение[1]. Этот документ значительно снижает барьеры для истцов по всей Европе, которые могут подавать иски об ответственности за вред, причиненный ИИ. Критики утверждают, что это лишь полумера, которая не позволяет в полной мере решить проблему вреда, связанного с ИИ[2].

Неправомерное использования генеративного ИИ и его ответственность за аудиодипфейки

Учитывая реальные последствия использования ИИ, суды, политики и ученые изучают, как существующие режимы ответственности могут взаимодействовать с ИИ, чтобы стимулировать его безопасное развитие и использование. Однако, как замечают *Бао Кхам Чау* (Корнеллский технологический институт (США)) и *Джордж Хэ* (Лаборатории инноваций Гарвардской библиотеки), эти режимы не учитывают должным образом сложность обучения базовых генеративных моделей ИИ, не принимают во внимание то,

[1] European Council. EU brings product liability rules in line with digital age and circular economy. – 2024. – URL: https://www.consilium.europa.eu/en/press/press-releases/2024/10/10/eu-brings-product-liability-rules-in-line-with-digital-age-and-circular-economy/.

[2] Hacker, P. The European AI liability directives – Critique of a half-hearted approach and lessons for the future, Computer Law & Security Review, 51 (2023), 105871. – 2023. – URL: https://doi.org/10.1016/j.clsr.2023.105871; https://ssrn.com/abstract=4279796

кому принадлежит инфраструктура для обработки данных, обучения моделей и их развертывания. Так, инструменты ИИ, позволяющие генерировать аудио, похожее на настоящее, становятся все более доступными и создают реальную угрозу для моделей, например монетизации существующих игроков музыкальной индустрии. Из-за сложной цепочки поставок данных трудно определить, кто несет ответственность за конкретное нарушение авторских прав. Такой подход к определению ответственного лица может привести к фатальным последствиям при распределении ответственности [1].

Существующие механизмы ответственности

По состоянию на декабрь 2024 г., отмечают Бао Кхам Чау и Дж. Хэ, в мире насчитывалось более 1600 политических инициатив, направленных на регулирование ИИ. В своей статье «Аудиодипфейки и регулирование со стороны “хозяев творчества”» они рассматривают три основные нормативно-правовые базы – американскую, европейскую и китайскую – в целях признать их глобальное влияние на развитие ИИ и управление им. В статье дается обзор этих баз и выявляются недостатки законодательства, особенно в том, что касается распределения ответственности за неправомерное использование генеративного ИИ. Анализ показал, что ни один из этих режимов не учитывает в достаточной мере то, что создатели базовых моделей, как их называют авторы – «хозяева творчества», – должны нести ответственность на разных этапах.

Так, в США разработка генеративного ИИ по большей части не регулируется. Несмотря на то что могут применяться такие механизмы, как судебная практика согласно Первой поправке и разд. 230 Закона о порядочности в сфере коммуникаций (the Communications Decency Act), они не обеспечивают всеобъемлющего нормативного регулирования. Более того, судья Горсач даже поставил под сомнение применимость ст. 230 к контенту, созданному ИИ (Gonzalez vs Google LLC, 2023) [1].

Всеобъемлющей федеральной системы ИИ не существует, утверждают Бао Кхам Чау и Дж. Хэ. Вместо этого несколько штатов, включая Калифорнию, Нью-Йорк и Иллинойс, приняли законы, касающиеся различных аспектов разработки и использования ИИ. Например, Калифорния приняла закон, требующий от своего Технологического департамента провести всестороннюю инвентаризацию всех высокорискованных вычислительных процессов,

полученных на основе машинного обучения, статистического моделирования, анализа данных, используемых в государственных учреждениях. Этот закон содержит перечень мер для снижения рисков или блокирования дискриминационных, предвзятых решений, принимаемых соответствующими вычислительными процессами (Cal. Gov't Code § 11,546.45.5, West). Аналогичным образом в штате Мичиган принят закон, требующий раскрытия информации о том, была ли политическая реклама полностью или по существу сгенерирована искусственным интеллектом (Mich. Comp. Laws Ann. § 169.247, West). В дополнение к этим законам штатов более 40 штатов внесли на рассмотрение более 400 законопроектов, связанных с ИИ [1].

Администрация Дональда Трампа в мае 2025 г. утвердила первый в истории федеральный закон, направленный против распространения «дипфейков» – поддельных изображений и видео, созданных с помощью нейросетей. Документ, получивший название Закона о борьбе с известными случаями эксплуатации путем блокировки технологических дипфейков на веб-сайтах и в сетях, или Закона о блокировке («Take It Down Act»)[1], вводит уголовное наказание за публикацию материалов без согласия человека, включая контент, сгенерированный ИИ[2].

В *КНР* активно разрабатываются нормативные акты в сфере ИИ. Как и в случае с американским и европейским подходами, некоторые китайские нормативные акты не направлены непосредственно на регулирование ИИ, но охватывают смежные области. Три из них имеют непосредственное отношение к генеративному ИИ и дипфейкам:

Положение об управлении алгоритмическими рекомендациями в информационных интернет-сервисах 2021 г. (*互联网信息服务算法推荐管理规定) (Положение о китайских алгоритмических рекомендациях*), которое в широком смысле распространяется на интернет-сервисы, использующие алгоритмические рекомендации, такие как социальные сети и электронная

[1] Подробнее об этом Законе см.: Take It Down Act. – URL: https://www.govtrack.us/congress/bills/119/s146/text; https://lawforeverything.com/take-it-down-act/; https://en.wikipedia.org/wiki/TAKE_IT_DOWN_Act?ysclid=mikd5mcirq752029893 (дата обращения: 15.11.2025).

[2] Дипфейки под запретом: в США начали криминализировать ИИ-подделки. – URL: https://vgtimes.ru/news/126535-dipfeyki-pod-zapretom-v-ssha-nachali-kriminalizovat-ii-poddelki.html (дата обращения: 18.10.2025).

коммерция. Оно предоставляет пользователям право отключать алгоритмические рекомендации, удалять теги персонализации и получать разъяснения о влиянии алгоритмов (ст. 17). Кроме того, был введен реестр алгоритмов, согласно которому поставщики должны предоставлять такую информацию, как: название поставщика, тип алгоритма, отчеты о самооценке и отображаемый контент (ст. 24). За нарушения предусмотрены штрафы в размере от 10 тыс. до 100 тыс. юаней (ст. 31);

Положение об управлении интернет-информационными сервисами глубокого синтеза 2021 г. (互联网信息服务深度合成管理规定) (З*акон о дипфейках*). Этот Закон требует, чтобы дипфейки были помечены соответствующим образом, чтобы их содержание не «вводило общественность в заблуждение» (ст. 17). Хотя в Законе о дипфейках в Китае прямо не указаны меры наказания за его нарушение, в нем говорится, что нарушители «будут наказаны в соответствии с действующими законами и административными постановлениями» (ст. 22);

Временные меры по управлению сервисами генеративного искусственного интеллекта 2023 г. (生成式人工智能服务管理暂行办法) (*Китайский регламент о генеративном искусственном интеллекте*). Этот Регламент распространяется на использование всех технологий генеративного ИИ, которые применяются для предоставления услуг населению, что, в частности, исключает разработку и применение технологий генеративного ИИ, которые не использовались для предоставления услуг населению (ст. 2). Китайский регламент о генеративном ИИ налагает весьма обременительные обязательства на поставщиков генеративного ИИ, требуя от поставщиков обеспечения того, чтобы права интеллектуальной собственности не нарушались, и чтобы поставщики «применяли эффективные меры для повышения качества обучающих данных и повышения достоверности, точности, объективности и разнообразия обучающих данных» (ст. 7). Если будет установлено, что поставщики генеративного ИИ нарушили китайское законодательство о генеративном ИИ, они могут быть привлечены к ответственности в соответствии с положениями законов КНР о кибербезопасности, о безопасности данных, о защите персональных данных, о научно-техническом прогрессе и других подобных законов и административных постановлений (ст. 21).

В Китае законодательство не предусматривает оптимального распределения ответственности между субъектами, участвующими в создании результатов работы генеративного ИИ. Например, хотя

китайский Закон о дипфейках требует, чтобы «поставщики услуг глубокого синтеза» наносили водяные знаки на результаты работы, это требование может быть возложено также на конечных потребителей базовых моделей (Закон о дипфейках, ст. 17, 23). Согласно китайскому определению, «поставщик услуг глубокого синтеза» – организация, которая дорабатывает (т.е. тонко настраивает) предварительно обученную базовую модель, способную генерировать материалы, нарушающие авторские права, – будет нести ответственность, даже если она не знала о таких материалах. Китайские суды уже возлагали ответственность на конечных потребителей генеративного ИИ в делах о нарушении авторских прав [1]. В связи с этим большое внимание в статье *Бао Кхам Чау* и *Дж. Хэ* уделяется вопросу ответственности владельцев креативных технологий ИИ, авторы рассматривают, какие стороны должны нести бремя ответственности таким образом, чтобы максимально стимулировать инновации и минимизировать ущерб. Они считают, что возлагать ответственность на арендаторов нецелесообразно, поскольку те не контролируют большие объемы данных для обучения базовой модели и алгоритмы генеративного ИИ. Всё в предварительно обученных базовых моделях контролируется арендодателями [ibid.].

Однако возлагать ответственность на владельцев креативных инструментов тоже проблематично, поскольку их модели могут быть использованы непредсказуемым образом. В самом простом случае владелец, который обучает базовую модель для злонамеренного использования, будет нести полную ответственность за такое использование ИИ. Однако сложность в том, что большинство базовых моделей не предназначены для конкретно злонамеренных действий или использования с высоким риском. Кроме того, если арендаторы (или субарендаторы) перепрофилируют (т.е. дорабатывают) базовые модели для другого законного применения, становится труднее определить, кто несет ответственность, если модели выдают вредоносные результаты.

Предлагаемая авторами система ответственности повышает прозрачность и ускоряет внедрение инноваций, если она (ответственность) по умолчанию возлагается непосредственно на арендодателей и тем самым стимулирует организации, обладающие наибольшим контролем, внедрять надежные меры безопасности и смягчения последствий. В отличие от общей системы ЕС, предусматривающей ответственность разработчиков систем ИИ без каких-либо исключений, китайская система предоставляет арендода-

телям четкие механизмы – такие, как журналы аудита и «красные команды» – для демонстрации ответственного поведения и надлежащего распределения ответственности в случае необходимости, расширяет существующие практики возмещения ущерба. Таким образом, китайский подход использует устоявшиеся правовые принципы для обеспечения более безопасной работы ИИ, не создавая при этом чрезмерных препятствий для инноваций [1].

Заключение

Анализ современных исследований показывает, что интеграция ИИ во все сферы общественной жизни должна быть сопряжена с жесткими и четкими правилами его разработки и использования, соблюдения фундаментальных прав человека и возложения ответственности в случае их нарушения. Труды некоторых участников Кембриджского форума «Искусственный интеллект: право и управление», отраженные в данном обзоре, подтверждают, что:

1) защита основных прав человека лежит в основе глобальных и академических проблем, связанных с технологиями ИИ. Достоинство, свобода, равенство, солидарность и справедливость являются краеугольным камнем человекоориентированного подхода к регулированию ИИ. Закон должен гарантировать индивидуальные меры защиты, предупреждать угрозы и риски, содержать меры предосторожности, направленные на обеспечение справедливого результата применения ИИ [3];

2) расширение возможности управления ИИ с помощью механизмов возмещения ущерба предполагает критически оценивать, в достаточной ли мере нынешнее управление ИИ удовлетворяет потребность в средствах правовой защиты от вреда, причиняемого ИИ, и возмещении ущерба. Обращение в суд, омбудсмену и другие способы правовой защиты позволяют людям, на которых повлияли системы ИИ, отстаивать свои права, особенно в тех случаях, когда нарушаются или подрываются принципы равенства, инклюзивности и справедливости [4];

3) негативные последствия дипфейков становятся все более очевидными по мере того, как расширяется применение генеративного ИИ. Его интеграция в различные приложения сопряжена со значительными рисками для национальной и личной безопасности; соответственно, создает новые проблемы для гражданского регулирования, например, закон об авторском праве. Учитывая реальные последствия, суды, политики и наука изучают, как суще-

ствующие режимы ответственности могут применяться к ИИ, чтобы стимулировать его безопасное развитие и использование. Однако, как показала практика, введенные меры ответственности не учитывают должным образом сложность обучения базовых генеративных моделей ИИ и их текущую экосистему. Поскольку компании-арендодатели владеют всей критически важной инфраструктурой и имеют технические возможности для контроля за ее использованием, предлагается обязать этих «арендодателей креативности» предоставлять технологии ИИ без дефектов и нести ответственность в случае неправильного использования генеративного ИИ [1].

Список литературы

1. Bao Kham Chau, George He. Audio deepfakes and the regulation of the landlords of creativity // Cambridge Forum on AI: Law and Governance. – 2025. – Vol. 1. – e30. – URL: https://doi.org/10.1017/cfl.2025.10012 (дата обращени: 17.10.2025).
2. Cambridge Forum on AI: Law and Governance. – 2025. – Vol. 1. – URL: https://www.cambridge.org/core/journals/cambridge-forum-on-ai-law-and-governance/article/individual-safeguards-in-the-era-of-ai-fair-algorithms-fair-regulation-fair-procedures (дата обращения: 17.10.2025).
3. Grozdanovski L., Cooman J. de. Individual safeguards in the era of AI: Fair algorithms, fair regulation, fair procedures // Cambridge Forum on AI: Law and Governance. – 2025. – Vol. 1. – e18. – URL: https://doi.org/10.1017/cfl.2025.10 (дата обращения: 17.10.2025).
4. Pi Y., Proctor M. Toward empowering AI governance with redress mechanisms // Cambridge Forum on AI: Law and Governance. – 2025. – Vol. 1. – e24. – URL: https://doi.org/10.1017/cfl.2025.9; https://www.cambridge.org/core/journals/cambridge-forum-on-ai-law-and-governance/article/toward-empowering-ai-governance-with-redress-mechanisms/A1EBCD6CAA146F503C8F6842914F3FB3 (дата обращения: 17.10.2025).