САМОЙЛЮК РОСТИСЛАВ НИКОЛАЕВИЧ[1]

roman7771777@yandex.ru

КАРПУНИНА ЕКАТЕРИНА СЕРГЕЕВНА[2]

ЯНБАРИСОВА РУМИЯ РАФАИЛОВНА[3]

[1,2,3] Всероссийский институт повышения квалификации сотрудников МВД России (Московская область, Россия)

Искусственный интеллект и права человека
(на примере Европейского союза)

Аннотация. В статье рассматриваются угрозы, создаваемые для прав человека быстрым развитием искусственного интеллекта[1], некоторые потенциальные международно-правовые меры по возможному противодействию таким угрозам, потенциальная переклассификация некоторых систем искусственного интеллекта, которые в настоящее время определяются как представляющие ограниченный риск, в системы, представляющие значительный риск или запрещенные системы.

Целью настоящей статьи является описание ключевых особенностей подхода Европейского союза[2] к регулированию искусственного интеллекта в контексте защиты прав человека, а также выявление как его достижений, так и недостатков, предложение улучшений существующих положений.

Ключевые слова и словосочетания: Европейский союз, право Европейского союза, искусственный интеллект, правовое регулирование искусственного интеллекта, европейский подход, права человека

***Для цитирования**: Самойлюк Р. Н., Карпунина Е. С., Янбарисова Р. Р. Искусственный интеллект и права человека (на примере Европейского союза) // Вестник ВИПК МВД России. – 2024. – № 4 (72). – С. 145-152 ; doi: 10.29039/2312-7937-2024-4-145-152*

KARPUNINA EKATERINA S.[2]
YANBARISOVA RUMIYA R.[3]

[1, 2, 3] Advanced Training Institute of the MIA of Russia (Moscow region, Russia)

ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS

[1]Далее – ИИ.
[2]Далее – ЕС.

(USING THE EXAMPLE OF THE EUROPEAN UNION)

Annotation. The article examines the threats posed to human rights by the rapid development of artificial intelligence, as well as some potential international legal measures to possibly counter such threats.

The article also discusses the potential reclassification of some artificial intelligence systems, which are currently defined as presenting limited risk, into systems that pose significant risk or prohibited systems.

Thus, the purpose of this article is to describe the key features of the European Union's approach to regulating artificial intelligence in the context of human rights protection, as well as to identify both its achievements and shortcomings, and propose improvements to existing provisions.

Key words and word combinations: European union, European union law, artificial intelligence, legal regulation and, European approach, human rights.

For citation: *Samoylyuk R. N., Karpunina E. S., Yanbarisova R. R. Artificial intelligence and human rights (using the example of the European union) // Vestnik Advanced Training Institute of the MIA of Russia. – 2024. – № 4 (72). – P. 145-152; doi: 10.29039/2312-7937-2024-4-145-152*

Активные усилия Европейского союза в области регулирования искусственного интеллекта кажутся особенно актуальными для исследований, учитывая подход, ориентированный на права граждан.

В настоящее время интенсивное развитие регулирования искусственного интеллекта в ЕС (компромиссный текст, представленный Советом ЕС, поправки Европейского комитета регионов, мнения заинтересованных сторон и правозащитных организаций и т.д.) делает исследование особенно своевременным из-за освещения проблемных аспектов.

13 марта 2024 г. Европейский парламент принял Регламент Европейского союза об искусственном интеллекте, устанавливающий гармонизированные правила в отношении искусственного интеллекта (далее – Регламент об искусственном интеллекте, Регламент) [1].

До настоящего времени он остается единственным всеобъемлющим документом, направленным на регулирование практически всех аспектов создания и применения технологий искусственного интеллекта. Само существование такого документа представляет собой новый этап в регулировании ИИ, об этом свидетельствуют следующие аспекты:

во-первых, это станет важным примером унификации правил на региональном уровне в области искусственного интеллекта;

во-вторых, из-за своего экстерриториального характера это окажет влияние на компании, расположенные за пределами ЕС, национальные законодательства третьих стран, а также международное право в целом.

Учитывая динамичное развитие индустрии искусственного интеллекта, а также разногласия по наиболее актуальным вопросам, таким как запрещенные приложения искусственного интеллекта, сертификация на основе самооценки, неудивительно, что утверждение и принятие Регламента рассматривалось более 3 лет. Более того, уместно также вспомнить, что, хотя проект Общего регламента по защите данных (GDPR)[1] был впервые опубликован в 2012 году, окончательная версия была принята только в 2016 году [2]. Это говорит о том, что окончательное принятие Регламента об искусственном интеллекте требовало временных затрат для всестороннего его изучения.

В контексте развития и широкого использования технологий искусственного интеллекта крайне важно уделять внимание сфере прав человека, на которую напрямую влияют системы искусственного интеллекта.

В этой связи, а также учитывая риски и влияние технологий искусственного интеллекта на права человека, необходимо определить возможные подходы к их минимизации. Таким образом, в контексте заявленных целей интересно рассмотреть подход ЕС к регулированию ИИ.

---

[1]*GDPR* – это регламент ЕС, который обновляет и расширяет более раннюю Директиву о защите данных (DPD), впервые принятую в 1995 году. Регламент GDPR посвящен обеспечению конфиденциальности данных физического лица, будь то клиент, сотрудник или деловой партнер. Цель GDPR заключается в обеспечении защиты персональных данных граждан ЕС, независимо от того, проживают они в ЕС или в другом месте.

Это особенно актуально с учетом динамичности среды регулирования и изменений в области искусственного интеллекта, что позволяет выделить спорные аспекты.

Так, например, опасения, высказываемые многими российскими и зарубежными экспертами по поводу чрезмерно ограничительного характера Регламента ЕС об искусственном интеллекте, при ближайшем рассмотрении обнаруживают, что многие деликатные аспекты остались нерешенными.

Например, классификация систем распознавания эмоций как представляющих ограниченный риск является спорной, более того, список исключений, предусмотренных в запрете на использование систем биометрической идентификации, довольно широк.

Основные результаты анализа регламента об искусственном интеллекте. После изучения основных рисков нарушений прав человека в свете разработки и внедрения технологий искусственного интеллекта в данном исследовании выявлены проблемные аспекты: отсутствие прозрачности, предвзятость, инвазивность и дискриминация систем искусственного интеллекта. Этот тезис позволяет сконцентрироваться на анализе Регламента об искусственном интеллекте и предлагаемых поправках, касающихся наиболее спорных аспектов.

Регламент об искусственном интеллекте содержит положения, направленные на снижение риска нарушений прав человека, связанных с развитием технологий искусственного интеллекта, а также обеспечение безопасного и «этичного» внедрения искусственного интеллекта.

Регламент предусматривает запрет определенных систем искусственного интеллекта, использование которых может существенно нарушать права человека. Сюда входят системы биометрической идентификации, используемые в общественных местах в правоохранительных целях, системы социальной оценки и ряд других. Учитывая, что использование таких систем может быть агрессивным и усиливать дискриминацию, их запрещение кажется вполне разумным.

Помимо определения систем высокого риска, Регламент об искусственном интеллекте включает в себя значительное количество требований, направленных на регулирование создания и внедрения этих систем. Например, требуется применять систему управления рисками для контроля рисков на протяжении всего жизненного цикла такой системы искусственного интеллекта. С этой целью Регламент об искусственном интеллекте формулирует требования к прозрачности системы и контролю со стороны человека.

Регламент также содержит положения для наборов данных, которые должны быть актуальными, репрезентативными, полными, безошибочными и обладать соответствующими статистическими характеристиками. Уменьшая количество искажений в наборах данных, это положение призвано уменьшить потенциальную предвзятость систем и, как следствие, количество дискриминационных решений.

Регламент также предусматривает требование к системам высокого риска проходить оценку соответствия; это призвано гарантировать, что только те системы, которые отвечают всем необходимым требованиям, являются безопасными, будут допущены к выходу на рынок ЕС.

Создание специализированного наднационального органа и национальных надзорных органов, предусмотренных Регламентом, призвано облегчить координацию в области искусственного интеллекта и обеспечить выполнение положений об искусственном интеллекте.

Регламент об искусственном интеллекте также содержит положения о значительных штрафах, налагаемых в случае нарушения требований Регламента.

Обзор предлагаемых изменений в Регламент об искусственном интеллекте, ссылающийся на компромиссный текст Совета ЕС, поправки Комитета регионов ЕС, совместную позицию

EDPB[1] и EDPS[2], выявляет несколько улучшений по сравнению с первоначальной версией проекта Регламента об искусственном интеллекте, которые приведены ниже.

Изменения, предложенные Советом ЕС в отношении определения систем искусственного интеллекта, обеспечивают основу для отличия технологий искусственного интеллекта от других информационных технологий. В компромиссном тексте не упоминается программное обеспечение как единственная форма систем искусственного интеллекта.

Хотя подход, основанный на оценке риска, который включает в себя четыре уровня риска, остается неизменным, включены разъяснения относительно ИИ общего назначения, к которым Регламент об ИИ не применяется.

Разъяснены запрещенные виды использования ИИ. Запрет на социальный скоринг[3] также был распространен на отдельных лиц, в то время как запрет на использование систем биометрической идентификации в общественных местах теперь включает использование систем правоохранительными органами и от их имени, что позволяет распространить запрет на тех, кто сотрудничает с правоохранительными органами.

Приложение III, которое содержит восемь областей применения ИИ с высоким риском, обновлено. Добавлены следующие подпункты: защита окружающей среды (ИИ, предназначенный для контроля выбросов и загрязнения) как часть пункта 2, в то время как системы ИИ, используемые для расчета страховых премий, андеррайтинга и оценки претензий, описаны в пункте 5 (d). Системы, предназначенные для криминальной аналитики, исключены из сферы применения ИИ правоохранительными органами (пункт 6 (g).

Европейский комитет регионов указывает на необходимость систематического уведомления физических лиц о том, что они взаимодействуют с системой, а также на необходимость такого уведомления в отношении систем высокого риска (это не предусмотрено Законом об ИИ).

Европейский совет по защите данных (EDPB) и Европейский надзорный орган по защите данных (EDPS) призывают уделять пристальное внимание системам распознавания эмоций, которые, по их мнению, должны быть запрещены, за исключением строго определенных случаев (в Законе об искусственном интеллекте такие системы перечислены среди систем с ограниченным риском).

EDPB и EDPS призывают адаптировать процедуру оценки соответствия, чтобы предварительная оценка всегда проводилась независимыми третьими сторонами в отношении систем высокого риска.

Отмечая очевидные улучшения по сравнению с первоначальной версией Закона об искусственном интеллекте необходимо отметить несколько деликатных аспектов, которые также требуют пристального внимания:

Закон об ИИ и предлагаемые поправки не предусматривают механизмов обновления запрещенных приложений ИИ (ст. 5) или систем с ограниченным риском (ст. 52). Для приложений ИИ с высоким риском возможность обновления приложений ограничена определенными областями. В совокупности это подразумевает негибкость регулирования с точки зрения неспособности своевременно реагировать на возникающие угрозы и обеспечивать законодательную релевантность быстро прогрессирующему развитию ИИ. Таким образом, становится необходимым предусмотреть механизмы обновления и соответствующие критерии.

Манипулирование и искажение поведения людей, а также выявление и эксплуатация уязвимости определенных категорий граждан, по-видимому, включают вредные практики, которые уже нарушают права человека и, следовательно, не требуют дополнительных критериев физического или психологического вреда, как указано в ст. 5.

---

[1]*European Data Protection Board* (Европейский совет по защите данных) – это децентрализованный независимый орган Европейского союза, целью которого является обеспечение последовательного применения Общего регламента по защите данных (GDPR) и содействие сотрудничеству между органами ЕС по защите данных.

[2]*European Data Protection Supervisor* – независимый надзорный орган, основной задачей которого является мониторинг и обеспечение соблюдения европейскими учреждениями и органами права на неприкосновенность частной жизни и защиту данных при обработке персональных данных и разработке новых политик.

[3]*Социальный скоринг* – это вид скоринга, который оценивает клиента по его социальным характеристикам и прогнозирует его поведение с помощью анализа его присутствия в социальных сетях.

Среди областей применения ИИ высокого риска необходимо предусмотреть использование ИИ в секторе здравоохранения.

Использование систем искусственного интеллекта для оценки риска совершения индивидом преступления или его повторения, а также для прогнозирования преступления или его повторения на основе профилирования или оценки личных качеств и других характеристик вызывает много вопросов. Поскольку Закон об искусственном интеллекте рассматривает такие системы как системы высокого риска, уместно классифицировать такие системы или, по крайней мере, определенные практики их применения как запрещенные.

Поскольку предлагаемые правила для систем искусственного интеллекта высокого риска абстрактны по своей природе, они требуют разработки практических инструкций, чтобы обеспечить их выполнение в каждом конкретном случае.

Необходимо уточнить концепции прозрачности для ст. 13 (ИИ высокого риска) и ст. 52 (системы ограниченного риска) и включить в статью 13 обязательное уведомление лица о взаимодействии с системой ИИ.

Процедура оценки соответствия должна быть адаптирована таким образом, чтобы исключить возможность самооценки, по крайней мере на начальном этапе.

Пристальное внимание следует уделять системам с ограниченным риском (перечисленным в разделе 52: системы распознавания эмоций, системы биометрической категоризации, глубокие подделки). Важно классифицировать некоторые из них как запрещенные (например, системы распознавания эмоций) или как высокорискованные, чтобы они подпадали под соответствующее регулирование.

Важно распространить на системы с ограниченным риском правило, касающееся права человека отказаться от взаимодействия с системой в пользу человека, если это необходимо для защиты его или ее прав.

Общий обзор возможных нарушений прав человека искусственным интеллектом. Хотя преимущества использования ИИ могут быть значительными, открывая широчайшие перспективы для будущего человечества, некоторые системы и приложения ИИ, тем не менее, сопряжены со значительными рисками нарушения основных прав граждан. С точки зрения прав человека большинство проблем, связанных с использованием ИИ и интеграцией технологий в человеческое общество, сводятся к рискам нарушения этих прав. В этом контексте мы сосредоточимся как на конкретно существующих ныне проблемах, так и на тех, которые могут возникнуть в ближайшем будущем, не говоря уже о долгосрочных рисках использования ИИ, которые могут представлять угрозу существованию человеческой цивилизации как таковой. Однако эффективные правила, позволяющие сдерживать текущие риски, могут помочь подготовить почву для противодействия долгосрочным угрозам.

Сегодня во всех национальных и международных правовых документах в области ИИ подчеркивается необходимость защиты прав человека.

Например, в поправках к Закону об искусственном интеллекте Европейский комитет регионов указал на защиту прав граждан как на одну из целей регулирования, тем самым подчеркнув его связь с Хартией основных прав ЕС [3, с. 72-80].

Проблемные аспекты могут касаться как самого ИИ, так и его сущности, а также особенностей его применения.

Так, Р. Родригес выделяет следующие проблемные аспекты ИИ:

отсутствие алгоритмической прозрачности;

проблемы, связанные с предвзятостью, несправедливостью и дискриминацией;

трудности в оспаривании решений систем ИИ;

неблагоприятное воздействие на рынок труда;

проблемы, связанные с конфиденциальностью информации и защитой данных.

Эти аспекты часто взаимосвязаны. Например, отсутствие прозрачности делает невозможным оспаривание соответствующих решений системы, в то время как предвзятость в наборах данных может привести к несправедливым и дискриминационным решениям.

Более того, все приведенные выше проблемы так или иначе приводят к нарушениям прав человека.

Действительно, если мы рассмотрим каждую из проблем, присущих индустрии искусственного интеллекта (системная предвзятость и дискриминация, непрозрачность алгоритмов, конфиденциальность, защита данных и ответственность за вред, причиняемый системами искусственного интеллекта), все это в конечном итоге сводится к рискам

нарушения прав человека. Такие риски, которые ни в коем случае не являются абстрактными, наиболее выражены в особо чувствительных областях: правосудии, здравоохранении, общественной безопасности, занятости, – где использование алгоритмов искусственного интеллекта может нанести ущерб правам человека, подразумевая необходимость защиты.

Например, из-за отсутствия необходимой прозрачности в алгоритмах искусственного интеллекта могут возникать ситуации, когда люди, чьи права затрагиваются действиями или решениями системы, не знают причины, по которой им было отказано в конкретной услуге, или почему в отношении них было принято определенное решение.

Прозрачность систем искусственного интеллекта в узком смысле означает способность понимать и объяснять решения системы. Системы искусственного интеллекта характеризуются значительной сложностью. Более того, глубокая нейронная сеть самостоятельно обучается генерировать «эффекты черного ящика», что означает невозможность идентифицировать и объяснить каждый этап процесса в форме, понятной людям [4, с. 1143-1185]. В результате такой непрозрачности действия и решения систем искусственного интеллекта часто становятся необъяснимыми и не поддающимися отслеживанию, что приводит к неспособности доказать несправедливость решений, принятых системой, и приводит к фактической неспособности граждан защитить свои собственные права.

Проблемы несправедливости, предвзятости и дискриминации также становятся острыми. Хотя такие явления могут быть сгруппированы вместе из-за их значительной взаимосвязи, предвзятость не всегда приводит к несправедливости или дискриминации, но может оставаться незамеченным отклонением от нормы, которое никак не влияет на решение системы.

Как правило, алгоритмическая предвзятость обусловлена предвзятостью в наборах данных, которая возникает с момента сбора данных и может быть объяснена как неправильной работой с наборами данных, так и историческими искажениями. Распознав такую предвзятость данных, алгоритмы могут затем выявить дополнительные различия, чтобы усилить ее, что приведет к принятию дискриминационных решений.

Таким образом, предвзятость алгоритмов определяется существующими в обществе предубеждениями, а также разнообразным составом групп, работающих с данными.

Чтобы проиллюстрировать вышеупомянутые проблемы, приведем пример одного из самых значительных скандалов в масштабах ЕС, связанных с доступом граждан к социальным пособиям в Нидерландах [5].

В 2014 году при поддержке Министерства социальных дел и занятости Нидерландов некоторые города начали использовать систему Risico Indicatie (SyRI), которая предназначена для выявления мошенничества в секторе социального обеспечения. В процессе расчета рисков для прогнозирования вероятности мошенничества со стороны получателей пособий эта система собирает и анализирует огромные объемы данных. Однако люди из слоев общества с более низкими доходами подвергались непропорционально высокой оценке, что приводило к дискриминации. Более того, потенциальные получатели пособий не имели возможности узнать, как система принимает решения. В 2020 году голландский суд постановил, что использование текущей версии SyRI незаконно из-за нарушения права на неприкосновенность частной жизни в смысле, описанном в Европейской конвенции по правам человека. Суд указал, что система не была прозрачной, собирала слишком много данных, цели сбора данных не были достаточно ясными и конкретными [6].

В деле итальянской компании по доставке продуктов питания Deliveroo[1] [7] суд установил, что алгоритм, используемый для ранжирования курьеров компании и определения приоритета сотрудников при доступе к удобным временным интервалам доставки, был дискриминационным. В данном случае не были учтены причины, по которым курьер не сообщил, что не сможет выйти на работу, что означает, что грань между прогулом и отсутствием по уважительным причинам не была проведена.

---

[1]*Deliveroo* – британская онлайн-компания по доставке еды, основанная Уиллом Шу и Грегом Орловски в 2013 году в Лондоне, Англия. Она работает в Соединенном Королевстве, Франции, Бельгии, Ирландии, Италии, Сингапуре, Гонконге, Объединенных Арабских Эмиратах, Кувейте и Катаре. Ранее она работала в Германии, Тайване, Испании, Нидерландах и Австралии.

Системы, отслеживающие сотрудников на рабочем месте, являются источником различных социальных проблем, поскольку позволяют записывать и анализировать все движения и операции, выполняемые человеком, включая его местоположение, экран рабочего стола, тон голоса и другие характеристики. В частности, различные гаджеты используются в рамках так называемых оздоровительных программ для сотрудников, передающих данные об их здоровье своим работодателям.

Хотя использование таких обширных наборов данных в сочетании с инструментами анализа может повысить производительность труда сотрудников, снизить риски для здоровья и безопасности, уменьшить вероятность несчастных случаев, необходимо учитывать, что эти системы программируются людьми и могут быть не лишены субъективных человеческих предубеждений. Более того, учитывая способность ИИ к самообучению и самоорганизации, существует риск того, что он может самостоятельно перепрограммировать критерии для достижения поставленных целей, что приведет к дискриминации. Более того, даже если данные анонимизированы, сам инвазивный процесс сбора данных нарушает конфиденциальность, выходя за рамки рабочих процессов.

В сфере трудовых отношений также можно проследить долгосрочные риски. В будущем широкое использование систем искусственного интеллекта, вероятно, приведет к значительным изменениям в требованиях к работникам, созданию новых типов рабочих мест, а также неравенству на «новом» рынке труда.

Вышеупомянутые системы могут эффективно заменить работников-людей, которые в настоящее время отвечают за управление персоналом. Это относится не только к специалистам по персоналу, но и к сотрудникам в других областях. Со временем ИИ может вытеснить людей из многих сфер деятельности. Так, М.Л. Энтин указывает, что внедрение ИИ в долгосрочной перспективе не увеличивает человеческие возможности, а вместо этого создает им альтернативу на рынке труда, не оставляя людям ничего, что можно было бы противопоставить.

Таким образом, масштабные возможности, демонстрируемые ИИ, влекут за собой столь же масштабные риски применения. Уже возникло множество ситуаций, в которых права человека нарушаются из-за использования искусственного интеллекта; их число, несомненно, увеличится в будущем. Поэтому необходимо уделять внимание как текущим рискам, так и прогнозу будущих, связанных с использованием ИИ, особенно в области прав человека, а также разработать соответствующую нормативную базу, направленную на минимизацию таких рисков.

Список источников.

1.https://ai.gov.ru/knowledgebase/dokumenty-po-razvitiyu-ii-v-drugikh-stranakh/2024_reglament_evropeyskogo_soyuza_ob_iskusstvennom_intellekte_ano_cifrovaya_ekonomika_/.

2. Энтин М.Л., Энтина Е.Г. В поисках партнерства: Россия и Европейский союз в 2020 - первой половине 2021 года // Зебра Э., 2021.

3. Предвзятость и дискриминация в ИИ: междисциплинарная перспектива / Х. Феррер, Т. Нюэнен, Дж. М. Сач [и др.] // журнал IEEE Technology and Society. 2021. № 40 (2) // https://doi.org/10.1109/MTS.2021.3056293.

4. Хакер П. Обучение искусственному интеллекту справедливости: существующие и новые стратегии против алгоритмической дискриминации в соответствии с законодательством ЕС // Обзор законодательства общего рынка. 2018. № 55 (4) // https://doi.org/10.54648/cola2018095.

5. Родригес Р. Юридические проблемы и права человека, связанные с ИИ: пробелы, вызовы и уязвимости // Журнал ответственных технологий, 2020. № 4. Ст. 100005 // https://doi.org/10.1016/j.jrt.2020.100005.

6. AI WATCH определяет искусственный интеллект. К оперативному определению и таксономии искусственного интеллекта / С. Самойли, М. Лопес Кобо, Э. Гомес [и др.] // Объединенный исследовательский центр, 2020 // https://publications.jrc.ec.europa.eu/repository/handle/JRC118163.

7. Сяпка А. Этические и юридические вызовы искусственного интеллекта: ответ ЕС на предвзятый и дискриминационный ИИ // SSRN, 2018 // https://doi.org/10.2139/ssrn.3408773.