# Механизмы защиты прав человека в эпоху искусственного интеллекта

**Нестерова Вероника Александровна**

*Студент (магистр)*

*Московский государственный университет имени М.В.Ломоносова,*
*Юридический факультет, Москва, Россия*

*Работа подготовлена с использованием системы КонсультантПлюс*

В эпоху развития искусственного интеллекта различные права человека подвергаются значительному риску их нарушения. Автором данной статьи рассмотрены основные проблемы защиты прав человека, связанные с повсеместным использованием технологий искусственного интеллекта (далее - ИИ), а также возможные механизмы и инструменты для их разрешения. В связи с развитием ИИ под угрозой находятся: право на равенство - для систем ИИ характерна проблема предвзятости, в результате которой в отношении людей может возникать дискриминация по признакам пола, расы, национальности; право на тайну переписки; право на получение достоверной информации; право на труд; право на свободу творчества и иные.

Важно отметить, что для предотвращения нарушений прав человека, а также их защиты в быстро развивающемся мире технологий необходим междисциплинарный, всесторонний, системный подход. Изменение регулирования должно осуществляться на всех уровнях общественных отношений, как в частном, так и в публичном секторе, а также охватывать собой совокупность научных дисциплин, способных создать единую базу для защиты прав человека с учетом новых технологий. На данный момент в силу непредсказуемости технологий, нерешенности проблемы “черного ящика” в системах ИИ, в связи с чем контроль над функционированием таких систем ограничен даже со стороны их разработчиков, необходимо действовать в рамках политики превенции - максимально предотвратить любые возможности для нарушения посредством технологий ИИ прав человека. В рамках данной политики автор работы считает наиболее релевантным использовать следующие механизмы: максимальное распространение информации о присутствии в той или иной системе ИИ-агента посредством направления пользователям соответствующих уведомлений; повсеместная маркировка контента, созданного при помощи ИИ.

Уведомление о наличии ИИ-агента: возможно закрепление обязанности разработчиков цифровых платформ устанавливать уведомления о встроенных ИИ-агентах. Законодательное регулирование такого оповещения может быть установлено по аналогии с уведомлением о сборе cookie файлов, законодательно урегулированным в рамках Европейского союза и США. В соответствии с позицией Роскомнадзора такие файлы относятся к персональным данным, на них распространяется действие Федерального Закона №152-ФЗ “О персональных данных” (далее - ФЗ “О персональных данных”). Представляется, что закрепление на законодательном уровне обязанности создателей цифровых платформ уведомлять граждан об использовании на таких платформах систем ИИ релевантно предусмотреть в рамках соответствующего положения в ФЗ “О персональных данных”[1], так как внутри чатов с пользователями ИИ-агенты осуществляют сбор информации, которая может подпадать под определение персональных данных. Например, если при отборе кандидатур на рабочие места лицам было направлено уведомление о применении систем ИИ в процессе такого отбора, то при отклонении кандидатуры лицо вправе обратиться к компании с запросом о предоставлении аргументации такого отказа с целью убедиться, что отказ обоснован

объективными причинами и не является следствием проявления предвзятости ИИ. Прецеденты такого дискриминационного отбора кандидатур уже известны на практике: дело о нарушении трудовых прав и дискриминации женщин в 2018 году - в новом механизме рекрутинга компании Amazon был обнаружен дискриминационный принцип отбора персонала: самообучаемая система Amazon сделала вывод о предпочтительности мужских кандидатур, вследствие чего компания была вынуждена прекратить использование алгоритмов ИИ для подбора персонала[2].

Важной проблемой является возможное внедрение ИИ в чаты пользователей в социальных сетях, что ставит под угрозу права граждан на конфиденциальность, тайну переписки и тайну личной жизни. Подобные прецеденты уже существуют на практике: в 2024 году компания Meta *(признана экстремистской организацией, её деятельность в Российской Федерации запрещена)* интегрировала ИИ, от использования которого пользователи не могут отказаться, в принадлежащие компании социальные сети. Стало известно, что такой ИИ-сервис появился в чате группы для мам детей-школьников, утверждая о наличии у него ребенка, который тоже учится в школе в Нью-Йорке[3]. Сообщалось, что ИИ-сервис присоединяется к диалогу в чате при его упоминании, или если в сообщении будет задан вопрос, на который никто не ответит в течение часа. Такое использование сервисов искусственного интеллекта поднимает вопрос защиты прав граждан, охраны их персональных данных и приватности в эру повсеместного внедрения и использования технологий ИИ, способного собирать и собирающего данные для своего последующего обучения.

Таким образом, уведомление об использовании различными сервисами ИИ-агентов может предотвратить нарушение прав человека. Так, в Мерах по маркировке контента, созданного или синтезированного искусственным интеллектом, вступивших в силу 1 сентября 2025 в Китае, устанавливается обязанность раскрытия использования искусственного интеллекта в приложениях. В Калифорнии 19 сентября 2024 года был принят комплексный закон, который вступит в силу с 1 января 2026 года, SB 942 или “Закон о прозрачности ИИ” (далее - Закон о прозрачности ИИ), который, среди иных обязательств, требует от поставщиков ИИ-генерируемого контента предлагать пользователям возможность включать “явные” уведомления о том, что контент создан ИИ.

Маркировка контента, созданного при помощи ИИ: данная мера защиты направлена в первую очередь на защиту прав граждан на достоверную информацию, защиту от распространения сведений, порочащих репутацию человека, что особенно актуально для политической сферы - возможно нарушение прав лиц, непосредственно участвующих в избирательной кампании, а также лиц, осуществляющих свое активное избирательное право - нарушается их право на честные выборы, без фальсификаций и введения в заблуждение относительно кандидатов на выборные должности. Например, в 2023 году один из кандидатов в президенты Турции снял свою кандидатуру с предвыборной гонки за несколько дней до ее завершения в связи с распространением в сети Интернет компрометирующих его фото и видео, изготовленных при помощи технологии “дипфейк”. Также, на предвыборных дебатах один из кандидатов продемонстрировал видео, предположительно являющееся фейком, созданное при помощи технологии ИИ, в котором утверждалось, что незаконные в Турции военные группировки заявили о поддержке его главного оппонента на выборах[4].

Регулирование маркировки контента может осуществляться в трех направлениях: добровольная маркировка ИИ контента; обязательная маркировка политического контента и (или) дипфейков, добровольная маркировка прочего ИИ контента; обязательная маркировка любого контента, созданного с использованием систем ИИ[5]. Автор данной статьи придерживается мнения, что введение обязательной маркировки любого контента, созданного с использованием систем ИИ, является наиболее эффективным способом предотвращения возможности нарушения прав человека.

Определенное регулирование данного вопроса уже существует в США и Китае. В марте 2025 года в Сенат США был представлен “Закон о защите происхождения контента и целостности от редактирования и дипфейков”, содержащий запрет на удаление или изменение водяных знаков, наносимых на ИИ-контент; обязательство разработчиков генеративных ИИ добавлять информацию о происхождении контента; возможность для создателей прикреплять идентификационные данные к своему контенту. Калифорнийский Закон о прозрачности ИИ, который, в отличие от ранее упомянутых, уже принят, содержит требование к поставщикам ИИ-генерируемого контента внедрять маркировку, а также предоставлять бесплатные инструменты для обнаружения ИИ контента; включать “скрытые” сведения о происхождении контента. Закон предусматривает ряд санкций за нарушение указанных требований с целью предотвращения выхода на рынок сервисов, которые потенциально могут нарушить права граждан[6]. Стоит отметить, что Foundation for Individual Rights and Expression выступает против законов о маркировке ИИ контента, указывая, что такие меры являются формой принуждения к выражению мнения и нарушают права на свободу выражения, закрепленные Первой поправкой Конституции США.

В Китае с 1 сентября 2025 года вступили в силу “Меры по маркировке контента, созданного или синтезированного искусственным интеллектом” (далее- Меры)[7]. В них предусмотрено использование двух видов маркировки: явная - обозначения, которые видны в контенте или встроены в интерфейс; скрытая - встроенные в данные файлов технические инструменты для обнаружения ИИ-контента, которые не могут считываться пользователями. В Мерах подробно описаны правила размещения маркировок, а также требование о сохранении маркировки в скопированных и скачанных файлах, соблюдение которого должно обеспечиваться провайдерами сервиса. В документе установлена обязанность платформ, распространяющих контент, осуществлять проверку на наличие скрытых маркировок, в случае обнаружения которых необходимо обязательно установить на материал видимые предупреждения о наличии ИИ. При отсутствии скрытых маркировок контент может быть помечен платформой как “вероятно создан ИИ” или “подозревается создание ИИ”[8].

Таким образом, новое время порождает новые вызовы для прав человека, в связи с чем необходим новый подход и механизмы для их защиты.

**Источники и литература**


[1] Федеральный закон "О персональных данных" от 27.07.2006 N 152-ФЗ // СПС КонсультантПлюс

[2] Харитонова Ю. С., Савина В. С., Паньини Ф. Предвзятость алгоритмов искусственного интеллекта: вопросы этики и права // Вестник Пермского университета. Юридические науки. 2021. Вып. 53. С. 488–515. DOI: 10.17072/1995-4190-2021-53-488-515. С. 496.

[3] Facebook parent Meta Platforms unveils new set of artificial intelligence systems | AP News [Электронный ресурс] URL: https://apnews.com/article/meta-ai-assistant-llama3-large-language-models-llm-229b386ebfbdc23f0e9245a68f7eb2d0

[4] Кандидат в президенты Турции Мухаррем Инче снялся с выборов из-за скандального видео - informburo.kz [Электронный ресурс] URL: https://informburo.kz/novosti/kandidat-v-prezidenty-turcii-muxarrem-ince-snyalsya-s-vyborov-iz-za-skandalnogo-video?ysclid=mf6rmvh4g9543151545

[5] AI Governance: регулирование и комплаенс ИИ-систем [Электронный ресурс] URL: https://t.me/vychislit_po_IP/5129

[6] Bill Text - SB-942 California AI Transparency Act. [Электронный ресурс] URL: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB942

[7] 关于印发《人工智能生成合成内容标识办法》的通知_中央网络安全和信息化委员会办公室 [Электронный ресурс] URL: https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm
[8] Там же.