**Информация Российской Федерации**

**на запрос Управления Верховного комиссара ООН**

**по правам человека в связи с резолюцией 42/15 Совета ООН по правам человека «Право на неприкосновенность частной жизни в цифровой век»**

*Reference NV:P.EN*

Российская Федерация, изучив запрос Управления Верховного комиссара ООН по правам человека, хотела бы сообщить следующее.

**1.**

a)  Интеллектуальные автоматизированные системы охватывают все новые сферы человеческой деятельности, становятся фактором, влияющим на социально-экономическое и политическое развитие на национальном и глобальном уровне. Технологии искусственного интеллекта (ИИ) призваны нести пользу людям, внося свой вклад в стимулирование инклюзивного роста, устойчивое развитие и благосостояние населения. ИИ должен разрабатываться таким образом, чтобы соблюдалось уважение к верховенству права, демократическим ценностям и правам человека.

На любом этапе функционирования ИИ должно быть предусмотрено вмешательство человека в работу системы (человекоцентричный подход). При разработке алгоритмов работы ИИ важно не допускать введения ограничений права на неприкосновенность частной жизни. Необходимо обеспечить прозрачность и ответственное раскрытие информации о системах ИИ, чтобы люди понимали результаты использования ИИ и могли управлять им (прозрачность и объяснимость). Системы ИИ должны функционировать надёжно, гарантированно и безопасно в течение всего жизненного цикла. Потенциальные риски должны отслеживаться на постоянной основе, а случае необходимости должны оперативно приниматься соответствующие меры управления.

Организации и отдельные лица, разрабатывающие, внедряющие или эксплуатирующие системы ИИ, должны нести ответственность за их надлежащее функционирование. При разработке международных документов в сфере ИИ акцент следует сделать на включении в их тексты основополагающего принципа – человек имеет право на любом этапе разработки, внедрения или применения ИИ отменять или ветировать любые действия или решения, принимаемые ИИ. Речь идет о приоритете государства в обеспечении этого права.

b)  Системы искусственного интеллекта и в целом технологии ИИ являются лишь инструментом реализации той или иной политики (подходов), в том числе в области обеспечения прав человека, включая вопросы неприкосновенности частной жизни. В этом смысле, рассматривая ИИ как масштабируемую технологию с высокой степенью воздействия на общество, на данном этапе целесообразно вести речь не о «помощи ИИ в защите прав человека», а об исключении либо снижении рисков воздействия ИИ применительно к правочеловеческой области (включая неприкосновенность частной жизни).

Область использования ИИ весьма разнообразна. Сегодня уже рассматриваются отдельные случаи применения ИИ в юридической практике и судопроизводстве. Однако, чтобы оценить положительные (прогрессивные) стороны ИИ в сфере прав человека, целесообразно сначала комплексно изучить риски, угрозы и опасности, которые могут быть прямо или косвенно сопряжены с широкомасштабным внедрением систем ИИ (в том числе в судах в части неприкосновенности частной жизни и др.).

Применение ИИ в вышеупомянутых областях должны рассматриваться не обобщенно и универсально, а исключительно с точки зрения сугубо прикладного подхода – конкретный класс систем с характеристиками и функциями в сфере применения (для решения определенных задач).

В условиях отсутствия необходимого пакета технических стандартов и согласованной методологии сертификации, а также принимая во внимание сегодняшний уровень изученности технологий на основе ИИ, на данном этапе можно говорить лишь об использовании систем ИИ для анализа статистики и другой информации, собранной для целей мониторинга соблюдения прав человека. В этой связи целесообразно сосредоточиться на выявлении рисков,

а уже потом на поисках позитивного «продвижения» (использования).

с)  Применение систем ИИ, ввиду повышения интенсивности обработки персональных данных, сопряжено с дополнительными рисками нарушения тайны частной жизни граждан. В связи с этим предлагается стимулировать выработку и закрепление разработчиками и производителями соответствующих систем этических норм в области обработки информации (включая персональные данные), добросовестного информирования об основных функциональных особенностях систем ИИ и робототехники, а также внедрение систем добровольной сертификации соответствия таким нормам. Следует разработать профессиональные стандарты для инженеров по ИИ, воплотив правозащитную ответственность по выбору технических решений при проектировании и эксплуатации систем ИИ. Рекомендуется также создать системы рассмотрения жалоб и предоставления средств правовой защиты для своевременного реагирования на жалобы и апелляции пользователей. Необходимо регулярно публиковать данные о периодичности подачи жалоб и просьб о предоставлении средств правовой защиты в отношении ситуаций, где имели место сбои системы ИИ, а также о видах и эффективности имеющихся средств правовой защиты.

d)  Пренебрежение человекоцентричным подходом закладывает основы для нарушений демократических принципов и ценностей. Системы ИИ могут способствовать дискриминации какой-либо одной группы людей в пользу другой в результате заданных алгоритмов программирования, а также прийти к необходимости (целесообразности с точки зрения заложенного в них алгоритма, развитого посредством самообучения) такой дискриминации на основе собственного решения. Следует обеспечить создание эффективных и легитимных механизмов правовой защиты прав человека, предотвращения любых форм дискриминации и неравенства в связи с разработкой и применением ИИ.

e)  Взаимосвязь может строиться на рискориентированном подходе и принципе предосторожности. В данном случае технологии ИИ выступают в том же качестве, что и любые другие технологии, в том числе критические. Применительно к вопросам ИИ и прав человека в большинстве случаев речь идет не об ограничении машинами свобод, а об установленных (или отсутствующих) режимах (правилах, нормах) применения таких технологий. Например, видеокамера, подключенная к системе на основе ИИ, сама по себе не нарушает права граждан – это лишь инструмент, применение которого регламентируется госорганами, в соответствии с предусмотренными процедурами. Таким образом, множество рисков в части использования ИИ может быть снято на законодательном уровне.

Вопросы, связанные с неравенством, манипуляцией, предубеждениями, риском деградации, защитой человеческого достоинства, потерей человеком автономности, должны рассматриваться применительно к конкретным регуляторным решениям (законодательная база в определенной юрисдикции или правовом поле), видам систем ИИ, среде применения, карте рисков.

**2.**

a)  В России в соответствии с частью 1 статьи 23 Конституции Российской Федерации каждый имеет право на неприкосновенность частной жизни, личную и семейную тайну, защиту своей чести и доброго имени.

Статьей 137 Уголовного кодекса Российской Федерации предусмотрена ответственность за нарушение неприкосновенности частной жизни.

Кроме того, частью 2 статьи 150 Гражданского кодекса Российской Федерации (ГК РФ) установлено, что нематериальные блага защищаются в соответствии с ГК РФ и другими законами в случаях и в порядке, ими предусмотренных, а также в тех случаях и пределах, в каких использование способов защиты гражданских прав (статья 12 ГК РФ) вытекает из существа нарушенного нематериального блага или личного неимущественного права и характера последствий этого нарушения.

Кроме того, на защиту неприкосновенности частной жизни, в том числе при использовании искусственного интеллекта, направлены федеральные законы от 27.07.2006 г. № 152-ФЗ «О персональных данных» (статья 2), от 27.07.2006 г. № 149-ФЗ «Об информации, информационных технологиях и о защите информации» (статья 3), от 28.12.2013 г. № 395-ФЭ «О Государственной автоматизированной информационной системе «ЭРА-ГЛОНАСС» (статья 3), от 24.04.2020 г. № 123-ФЭ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации – городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона «О персональных данных».

В соответствии с Национальной стратегией развития искусственного интеллекта до 2030 года (утверждена Указом Президента Российской Федерации от 10.10.2019 г. № 490) одним из основных принципов развития и использования ИИ является защита прав и свобод человека: обеспечение защиты гарантированных российским и международным законодательством прав и свобод человека, в том числе права на труд, и предоставление гражданам возможности получать знания и приобретать навыки для успешной адаптации к условиям цифровой экономики (подпункт «а» пункта 19).

Согласно документу, к 2030 году в Российской Федерации должна функционировать гибкая система нормативно-правового регулирования в области ИИ, в том числе гарантирующая безопасность населения и направленная на стимулирование развития технологий ИИ, в том числе должны быть разработаны этические правила взаимодействия человека с ИИ.

Основными принципами развития и использования технологий искусственного интеллекта, соблюдение которых обязательно при реализации Стратегии, являются:

а) защита прав и свобод человека: обеспечение защиты гарантированных российским и международным законодательством прав и свобод человека, в том числе права на труд, и предоставление гражданам возможности получать знания и приобретать навыки для успешной адаптации к условиям цифровой экономики;

б) безопасность: недопустимость использования ИИ в целях умышленного причинения вреда гражданам и юридическим лицам, а также предупреждение и минимизация рисков возникновения негативных последствий использования технологий искусственного интеллекта;

в) прозрачность: объяснимость работы искусственного интеллекта и процесса достижения им результатов, недискриминационный доступ пользователей продуктов, которые созданы с использованием технологий искусственного интеллекта, к информации о применяемых в этих продуктах алгоритмах работы искусственного интеллекта;

г) технологический суверенитет: обеспечение необходимого уровня самостоятельности Российской Федерации в области искусственного интеллекта, в том числе посредством преимущественного использования отечественных технологий искусственного интеллекта и технологических решений, разработанных на основе искусственного интеллекта;

д) целостность инновационного цикла: обеспечение тесного взаимодействия научных исследований и разработок в области искусственного интеллекта с реальным сектором экономики;

е) разумная бережливость: осуществление и адаптация в приоритетном порядке существующих мер, направленных на реализацию государственной политики в научно-технической и других областях;

ж) поддержка конкуренции: развитие рыночных отношений и недопустимость действий, направленных на ограничение конкуренции между российскими организациями, осуществляющими деятельность в области искусственного интеллекта (пункт 19 Стратегии).

Следует отметить, что Российская Федерация является участником заключенной в рамках Совета Европы Конвенции о защите физических лиц при автоматизированной обработке персональных данных 1981 года, в соответствии с которой каждая Сторона обязана предусмотреть надлежащие санкции и средства правовой защиты на случай нарушения норм внутреннего законодательства, воплощающих основополагающие принципы защиты данных, изложенные в Главе II данной Конвенции.

b)  В разработке регулирующих документов (режимов) целесообразно двигаться от общего к частному, от универсального к предметному. Учитывая сложную природу систем ИИ и правочеловеческой проблематики, основой для разработки нормативно-правовых актов должны служить междисциплинарные научные исследования, не фрагментарного, а концептуального характера. Эта работа должна вестись с пониманием всего генезиса явлений и процессов, начиная от анализа проблемы на концептуальном, философском уровне и заканчивая практическими измерениями на основе согласованных метрик и методик (стандартов).

с)  Концепция развития регулирования отношений в сфере технологий ИИ и робототехники до 2024 года (далее – Концепция), утвержденная распоряжением Правительства Российской Федерации от 19.08.2020 г. № 2129-р, ставит во главу угла человека, его права, интересы, безопасность и благополучие. Документ не содержит положений о наделении какими-либо правами роботов или носителей ИИ. Человеко-ориентированный подход предполагает абсолютный запрет на причинение какого-либо вреда человеку по инициативе систем ИИ, фиксирует их подконтрольность человеку, проектируемое соответствие закону.

**3.**

а)  По мнению деловых кругов, российский рынок ИИ слабо отрегулирован, что создает проблемы для граждан. Соответствующие письма с предложением регламентировать рынок ИИ направлялись в Совет Федерации, Госдуму, Минцифры, Роскомнадзор. Риски дискриминации ИИ граждан и организаций возникают из-за возможного наличия предубеждений в алгоритмах, умышленно или бессознательно привнесенных разработчиками, некачественных или неполных данных, используемых для обучения системы или принятия решений в конкретной ситуации. При этом выявить ошибки ИИ и обжаловать их сейчас крайне затруднительно, из-за того что доступ сторонних экспертов к ИИ-системам закрыт.

Например, в сфере госуправления системы алгоритмического принятия решений используются для привлечения к административной ответственности на основе данных видео- или фотофиксации, а также на основе иных технических данных. Зачастую ИИ отклоняет заявки на кредиты, поданные вполне добросовестными заемщиками, на основании неверно интерпретированных данных и сырых алгоритмов. При решении этого вопроса человеком оно могло бы быть иным.

В соответствии с вышеупомянутой Концепцией, следует поддерживать развитие регулирования, вырабатываемого и приводимого в исполнение силами участников рынка (саморегулирование), включая принятие и использование документов национальной системы стандартизации, кодексов (сводов) этических правил и иных документов саморегулируемых организаций, а также иных инструментов.

Наглядным примером может служить деятельность ПАО Сбербанк, одной из первых компаний в России, утвердившая корпоративные этические принципы разработки и применения ИИ. Данные принципы устанавливают применение ИИ с соблюдением законодательства, в том числе требований конфиденциальности, и уважением к частной жизни человека и коммерческой тайне, и фиксируют, в том числе: контролируемость и управляемость систем ИИ, прозрачность и предсказуемость функционирования технологий ИИ, стабильность и надёжность систем ИИ, ответственное применение ИИ, непредвзятость ИИ.

b)  Контроль алгоритмов ИИ в обязательном порядке должен закрепляться за государствами. Важно не допускать негативного влияния алгоритмических процессов на сознание человека, на формирование поведенческих моделей и потребительских запросов. Необходимо выработать глобальные принципы по социальной ответственности компаний и организаций, специализирующихся на вопросах ИИ. Компаниям и организациям следует ориентировать свои стандарты, правила и структуру системы на универсальные принципы прав человека. Эффективная работа по регулированию ИИ требует формирования позитивной повестки дня в обществе и на международном уровне.

Пункты с) и d) могут быть решены за счет внедрения традиционных инструментов оценки. Например, для оценки воздействия на права человека (human rights impact assessment) существует множество практических руководств и различных критерий, в том числе в ООН.

Вопросы инженерной и технологической защиты могут быть решены в рамках стандартизации и последующей сертификации. При этом следует учитывать, что система правил не работает идеально. Существуют риски, связанные с использованием конкурентных преимуществ одной страны или группы государств, которые могут получить право интерпретации и трактовки, разработанных ими технических норм. Иными словами «кто придумывает правила, тот ими выгодно пользуется». Тем не менее, любое продвижение в сфере стандартизации и методик сертификации скорее приносит пользу, чем вред. Что касается разработки специализированных моделей управления данными и программным обеспечением, которые будет работать «во благо общества», то сначала следует завершить этап выявления, исключения и минимизации рисков «технологии не во вред». Пункты с) и d) носят во многом практический инженерный характер и лежат скорее в области компетенции технологов, программистов, разработчиков.