



Генеральная Ассамблея

Distr.: General
29 August 2018
Russian
Original: English

Семьдесят третья сессия

Пункт 74 b) предварительной повестки дня*

Поощрение и защита прав человека: вопросы прав человека, включая альтернативные подходы в деле содействия эффективному осуществлению прав человека и основных свобод

Поощрение и защита права на свободу мнений и их свободное выражение**

Записка Генерального секретаря

Генеральный секретарь имеет честь препроводить Генеральной Ассамблее доклад Специального докладчика по вопросу о поощрении и защите права на свободу мнений и их свободное выражение Дэвида Кея, представленный в соответствии с резолюцией 34/18 Совета по правам человека. В настоящем докладе Специальный докладчик рассматривает последствия применения технологий искусственного интеллекта для осуществления прав человека в информационной среде, уделяя особое внимание, в частности, праву на свободу мнений и их свободное выражение, неприкосновенность частной жизни и недискриминацию.

* A/73/150.

** Настоящий доклад был представлен с опозданием в связи с необходимостью отразить в нем последние события.



Доклад Специального докладчика по вопросу о поощрении и защите права на свободу мнений и их свободное выражение

Содержание

	<i>Стр.</i>
I. Введение	3
II. Понимание искусственного интеллекта	3
A. Что такое искусственный интеллект?	3
B. Искусственный интеллект и информационная среда	7
III. Правовая основа в области прав человека для искусственного интеллекта	12
A. Объем обязательств в области прав человека в контексте искусственного интеллекта	12
B. Право на свободу мнений	13
C. Право на свободу выражения мнений	15
D. Право на неприкосновенность частной жизни	16
E. Обязательства, касающиеся недискриминации	17
F. Право на эффективное восстановление в правах	19
G. Законодательные, нормативные и политические меры реагирования на появление искусственного интеллекта	20
IV. Правовая основа в области прав человека для искусственного интеллекта	21
A. Основные стандарты для систем искусственного интеллекта	22
B. Процессы для систем искусственного интеллекта	24
V. Выводы и рекомендации	26

I. Введение

1. Искусственный интеллект во все большей степени оказывает влияние на глобальную информационную среду. Он позволяет компаниям курировать результаты поиска и формировать новостные потоки, определять, что и когда видят пользователи. Технологии искусственного интеллекта используются компаниями-операторами социальных сетей в целях модерации содержания на платформах и часто представляют собой первую линию обороны в борьбе с материалами, которые могут нарушать действующие правила. Искусственный интеллект рекомендует пользователям ставить лайки или выступать в роли фолловеров, новостные сообщения для прочтения или места для посещения, рестораны, магазины и гостиницы. Он работает быстро, эффективно и масштабно, помогая крупнейшим компаниям в секторе информационно-коммуникационных технологий управлять огромные объемы контента, которые каждый день загружаются на их платформы. Технологии искусственного интеллекта обладают потенциалом для более широкого и оперативного обмена информацией и идеями в глобальном масштабе, открывая широкие возможности для свободы выражения мнений и доступа к информации. В то же время непрозрачность искусственного интеллекта создает риски для принятия самостоятельных решений, процесса именуемого в настоящем докладе «индивидуальной автономией и агентностью»¹. В этом контексте все поборники прав человека и защитники принципа верховенства права должны найти ответ на вопрос, имеющий глобальное значение: как государства, компании и гражданское общество могут обеспечить, чтобы применение технологий искусственного интеллекта служило делу укрепления и соблюдения права человека, а не приводило к ослаблению и ущемлению этих прав?

2. Настоящий доклад не претендует на то, что он является последним словом в области искусственного интеллекта и прав человека. Скорее, в нем предпринята попытка решить три задачи: определить ключевые термины, которые имеют важное значение для обсуждения вопросов искусственного интеллекта с точки зрения прав человека; установить правовую основу в области прав человека для искусственного интеллекта и представить некоторые предварительные рекомендации с тем, чтобы по мере дальнейшей эволюции технологий искусственного интеллекта обеспечить учет соображений, связанных с осуществлением прав человека. Настоящий доклад следует рассматривать в качестве дополнения к моему последнему докладу Совету по правам человека (A/HRC/38/35), в котором был представлен правозащитный подход к модерации онлайн-контента².

II. Понимание искусственного интеллекта

A. Что такое искусственный интеллект?

3. Термин «искусственный интеллект» часто используется для краткого описания системы оперативного принятия автономных и автоматизированных

¹ См. Mariarosaria Taddeo and Luciano Floridi, “How AI can be a force for good”, *Science*, vol. 361, No. 6404 (24 August 2018). URL: <http://science.sciencemag.org/content/361/6404/751.full>.

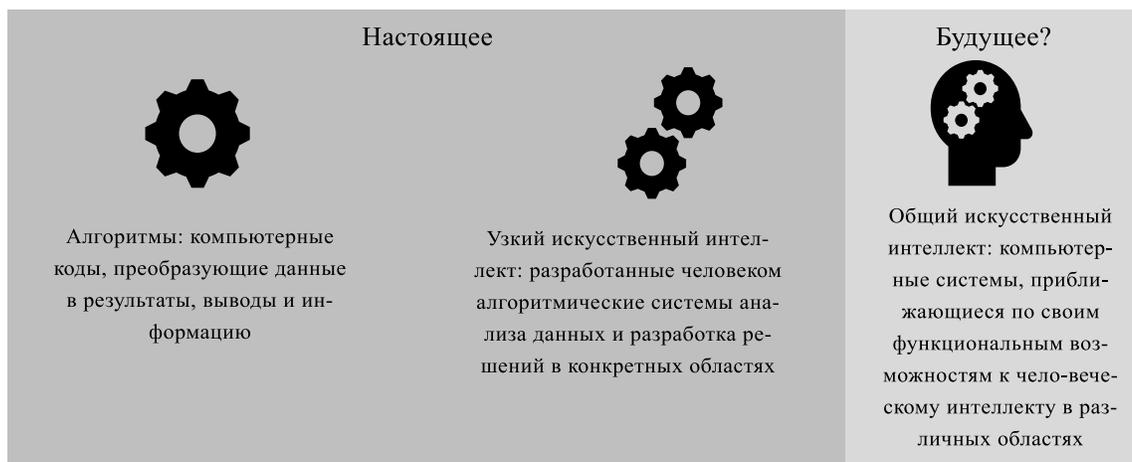
² В настоящем докладе были учтены результаты консультаций экспертов, проведенных в Женеве в июне 2018 года за счет дотации Европейского союза, а также соображения экспертов, высказанные в рамках разработки документа A/HRC/35/38 в 2017 и 2018 годах. Я хотел бы особо поблагодарить Карли Нист и Эймос Тох, предоставивших основные исследовательские материалы и участвовавших в подготовке этого проекта.

решений при помощи вычислительной техники. Искусственный интеллект не представляет собой какой-то единый механизм, но скорее комбинацию процессов и технологий, которые позволяют компьютерам, например, в рамках принятия решений и решения проблем, дополнять или заменять конкретные функции, которые в противном случае должны были бы выполнять люди³. В силу своего внутреннего содержания термин «искусственный интеллект» может создать ложное представление о том, что машинный интеллект функционирует по тем же принципам и правилам, что и мозг человека. Это не так. В широком смысле искусственный интеллект представляет собой систему, которая позволяет оптимизировать решение поставленных людьми вычислительных задач посредством многократных повторений и попыток. Однако этот термин закрепился в культуре, получил распространение в компаниях и правительстве, и я буду использовать его в этом докладе.

4. Среди общественности достаточно широко распространено мнение о том, что общество развивается в направлении появления общей формы искусственного интеллекта, представляющего собой все еще существующего в отдаленной перспективе потенциал («сингулярность»), который обеспечит компьютерным системам функциональные возможности, равные или даже превышающие возможности человеческого разума в различных областях⁴. Однако в ближайшей перспективе прогресс будет идти по линии создания форм искусственного интеллекта с более ограниченными функциональными возможностями, позволяющими компьютерным системам выполнять программные задачи в конкретных областях (на основе разработанных человеком алгоритмов). Так, например, узкий искусственный интеллект используется виртуальными собеседниками в голосовых помощниках мобильных устройств и службах по работе с клиентами, в онлайн-программах письменного перевода, в беспилотных автомобилях, в поисковых системах и картографических сервисах. Машинное обучение также использует узкий искусственный интеллект в целях обучения алгоритмов использованию баз данных для идентификации и решения проблем. Еще одним примером использования искусственного интеллекта являются «умные» домашние устройства, которые используют данные из повседневной речи и речевые модели для постоянного «обучения» в целях повышения эффективности обработки информации и более четкого реагирования на поступающие от пользователей вопросы. Во всех этих случаях важнейшую роль в разработке и распространении систем искусственного интеллекта играют люди, которые определяют параметры интеллектуальных приложений и, в зависимости от типа приложения, отбирают и маркируют базы данных и классифицируют результаты. Люди всегда определяют функциональные характеристики интеллектуальных приложений и используют полученные с их помощью результаты, в том числе, для определения того, в какой степени они дополняют или заменяют людей в процесс принятия решений.

³ См. AI Now, “The AI now report: the social and economic implications of artificial intelligence technologies in the near term”, 2016. URL: https://ainowinstitute.org/AI_Now_2016_Report.pdf; Специальный комитет Палаты лордов по вопросам искусственного интеллекта Соединенного Королевства Select Committee on Artificial Intelligence, AI in the United Kingdom: ready, willing and able?, 2018 («ИИ в Соединенном Королевстве: готов, желает и способен?»).

⁴ Article 19 and Privacy International, “Privacy and freedom of expression in an age of artificial intelligence”, London, 2018.



5. Основу искусственного интеллекта составляют разработанные и написанные людьми алгоритмы и компьютерные коды, содержащие инструкции о преобразовании данных в выводы, информацию или результаты. Алгоритмы прочно вошли в повседневную жизнь и играют важную роль в работе систем связи и объектов инфраструктуры. Огромные объемы данных, которые генерирует современная жизнь, и наличие потенциала для их анализа служат стимулом для развития искусственного интеллекта. Нет сомнения в том, что частный сектор рассматривает данные именно в этом ракурсе, полагая, что большие объемы данных более высокого качества помогают получать все более мощные и точные алгоритмы. Алгоритмические системы способны оперативно анализировать огромные объемы данных, позволяя интеллектуальным программам выполнять связанные с принятием решений функции, которые входили ранее в круг обязанностей людей, не оснащенных компьютерными средствами.



Хотя человек является неотъемлемой частью системы искусственного интеллекта, ряд характерных особенностей системы заставляют подвергнуть ее тщательному анализу с точки зрения соблюдения прав человека и изучить по крайней мере три аспекта ее функционирования, связанных с автоматизацией, анализом данных и адаптируемостью⁵.

6. **Автоматизация.** Автоматизация позволяет устранить участие человека в ряде сегментов процесса принятия решений и обеспечить решение конкретных задач при помощи вычислительной техники. Однако такой подход может давать положительные результаты для осуществления прав человека, только если проектное решение позволяет устранить элемент предвзятости со стороны человека. Например, автоматизированная система пограничного контроля может использовать объективные параметры, такие как информация об уголовном прошлом или визовом статусе, в целях идентификации подлежащих проверке лиц и ограничивать использование субъективных (и определяемых предубеждениями) оценок, основанных на таких признаках как внешние данные, этническая принадлежность, возраст или религия. Автоматизация также позволяет обрабатывать огромные объемы данных с недоступной человеку скоростью, что в потенциале может служить интересам охраны общественного порядка, здравоохранения и национальной безопасности. Однако автоматизированные системы используют наборы данных, которые с точки зрения разработки концепции и ее

⁵ Council of Europe, *Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications*, Council of Europe study, No. DGI (2017) 12, 2018. URL: <https://www.coe.int/en/web/freedom-expression/-/algorithms-and-human-rights-a-new-study-has-been-published>.

реализации оставляют место для предвзятости и, таким образом, могут приводить к дискриминационным последствиям. Например, упомянутая выше информация об уголовном прошлом или визовых данных, может содержать элементы предвзятости. Чрезмерная опора на автоматизированные решения и вера в их непогрешимость, а также отказ от признания этого принципиального соображения, могут отрицательно сказаться на анализе результатов, полученных с помощью искусственного интеллекта, и лишить доступа к средствам правовой защиты при оспаривании неблагоприятных решений, полученных с помощью искусственного интеллекта. Автоматизация может препятствовать обеспечению транспарентности и проведению изучения процесса, лишая даже благонамеренные власти возможности представить разъяснения по поводу полученных результатов⁶.

7. **Анализ данных.** Большинство приложений искусственного интеллекта построены на принципе обработки больших объемов данных. В основу системы искусственного интеллекта может быть положен любой набор данных, от данных о характерных привычках навигации в Интернете до данных о транспортных потоках на автомобильных дорогах. Некоторые наборы данных содержат информацию личного характера, в то время как многие другие включают анонимизированные данные. Использование искусственным интеллектом таких наборов данных вызывает серьезную озабоченность, в том числе в связи с возникновением вопросов относительно их происхождении, достоверности и прав граждан на эту информацию, способности систем искусственного интеллекта к идентификации обезличенных данных и предубеждений, которые могут быть встроены или привнесены в наборы данных через формы обучения или способы маркировки. Оценка данных с помощью искусственного интеллекта может способствовать установлению взаимосвязей, однако не обязательно причинной зависимости, что может привести к предвзятости и ошибочным результатам, которые трудно поддаются критическому анализу.

8. **Адаптируемость.** Системы машинного обучения, использующие искусственный интеллект, характеризуются адаптируемостью, поскольку лежащие в основе алгоритмы могут постепенно выявлять новые проблемы и разрабатывать новые решения. В зависимости от уровня контроля, системы могут выявлять закономерности и делать выводы, которые не могут быть получены людьми, писавшими программы и формулировавшими задачи. В то время, как, с одной стороны, такое отсутствие предсказуемости вселяет реальную надежду на то, что искусственный интеллект в перспективе может стать поистине трансформационной технологией, с другой стороны, это указывает на наличие в системе рисков, обусловленных тем, что исключение все большего числа людей из процесса постановки целей и оценки результатов функционирования системы искусственного интеллекта значительно усложняет задачу обеспечения транспарентности, подотчетности и доступа к эффективным средствам правовой защиты, равно как и прогнозирования и смягчения неблагоприятного последствий с точки зрения прав человека.

В. Искусственный интеллект и информационная среда

9. Использование искусственного интеллекта приводит к важным, а в некоторых случаях проблематичным последствиям для информационной среды, сложных технологических экосистем, платформ, частных и государственных субъектов, обеспечивающих доступ к информации и ее распространение с помощью цифровых средств. Алгоритмы и интеллектуальные приложения можно

⁶ Council of Europe, *Algorithms and Human Rights*.

обнаружить на всем пространстве Интернета, в цифровых устройствах, в технических системах, в поисковых программах, на платформах социальных сетей, в приложениях для обмена текстовыми сообщениями и механизмах информирования общественности. В соответствии с темой моего мандата, я хотел бы отметить следующие три сферы применения искусственного интеллекта в информационной среде, которые вызывают озабоченность:

10. **Представление контента и персонализация.** Социальные сети и поисковые платформы во все большей степени определяют способы получения доступа к информации, обмена идеями и распространения новостей. Алгоритмы и интеллектуальные приложения определяют масштаб, периодичность и целевую аудиторию для распространяемого контента. Огромные массивы данных, которые включают информацию об истории работы в Интернете, демографические данные о пользователях, семантический и эмоциональный анализ пользователей и многие другие элементы, используются для построения все более персонализированных алгоритмических моделей для приоритизации и курирования информации в целях представления определенной информации или сокрытия ее от пользователей. Платный, спонсируемый или помеченный хэштегом контент может продвигаться за счет исключения или понижения уровня приоритетности другого содержания. Новостные потоки социальных сетей подают информацию на основе субъективных представлений о том, насколько конкретный контент представляет интерес для пользователя и может удерживать его на сайте; в результате пользователи на своих платформах могут получать ограниченное освещение или не получать никакой информации об определенных важнейших общественных или политических событиях⁷. Искусственный интеллект формирует информационное пространство непрозрачным для пользователя способом и нередко принципы курирования информации неизвестны даже оператору информационной платформы.

11. Онлайн-поиск является одной из наиболее распространенных форм представления и персонализации контента с использованием искусственного интеллекта. Поисковые системы представляют результаты поиска (и дополняют или предсказывают запросы) с использованием систем искусственного интеллекта, обрабатывающих обширные объемы данных об индивидуальных и групповых пользователях. Поскольку контент с низким уровнем приоритетности и полностью исключенный из результатов поиска вряд ли будет представлен на сайте, поисковые приложения искусственного интеллекта оказывают значительное влияние на распространение знаний⁸. Аналогичным образом агрегаторы контента и новостные сайты⁹ отбирают информацию для пользователей не по признаку актуальности или важности событий, но по рекомендации приложений искусственного интеллекта, которые прогнозируют интересы и новостные предпочтения пользователей на основе обширных объемов данных. Следовательно, искусственный интеллект играет важную, но, как правило, скрытую роль в подборе информации для пользователей и в определении даже тех сведений, которые они будут воспринимать осознанно.

⁷ World Wide Web Foundation, “The invisible curation of content: Facebook’s News Feed and our information diets”, 22 April 2018. URL: <https://webfoundation.org/research/the-invisible-curation-of-content-facebooks-news-feed-and-our-information-diets/>.

⁸ Council of Europe, *Algorithms and Human Rights*.

⁹ См. например, “How Reuters’s revolutionary AI system gathers global news,” MIT Technology Review, 27 November 2017. URL: <https://www.technologyreview.com/s/609558/how-reuterss-revolutionary-ai-system-gathers-global-news/>; Paul Armstrong and Yue Wang, “China’s \$11 billion news aggregator Jinri Toutiao is no fake,” *Forbes*, 26 May 2017. URL: <https://www.forbes.com/sites/ywang/2017/05/26/jinri-toutiao-how-chinas-11-billion-news-aggregator-is-no-fake/#1d8b97804d8a>.

12. Использование искусственного интеллекта в целях представления контента идет в направлении все большей персонализации работы пользователей в режиме онлайн; в эпоху информационного изобилия¹⁰ персонализация может привести упорядоченность в хаотическую подачу данных в Интернете, позволяя пользователям находить запрашиваемую информацию. В числе выгод можно назвать возможность получения доступа к информации и услугам на большем количестве языков¹¹ или актуальной и важной информации в точки зрения личного опыта и предпочтений. Персонализация с использованием искусственного интеллекта может также ограничивать возможности ознакомления с другими точками зрения, что мешает индивидууму получать доступ к идеям и мнениям и обмениваться ими, преодолевая идеологические, политические и общественные разногласия. Такая персонализация может приводить к усилению предвзятости и стимулировать пропаганду и распространение материалов подстрекательского содержания или поощрять дезинформацию в целях удержания пользователей в режиме онлайн¹². Разумеется, для отдельных лиц причиной ограничения доступа к информации могут служить различные социальные и культурные контексты. Однако персонализация с помощью искусственного интеллекта посредством оптимизации методов удержания и стимулирования виральности в необходимых масштабах может ограничивать возможности индивидуального выбора определенных видов контента. Это особенно верно в силу того, что алгоритмы обычно понижают уровень приоритетности контента с более низким потенциалом удержания, в результате чего контент из независимых источников и генерируемый пользователями отодвигается на второй план¹³. Опытные операторы могут использовать основанные на правилах системы искусственного интеллекта, оптимизированные для удержания пользователя, в целях повышения уровня воздействия и, посредством присвоения популярных хэштегов или использования ботов, могуткратно увеличивать размер онлайн-аудитории в ущерб информационному разнообразию.

13. **Модерация и удаление контента.** Искусственный интеллект помогает компаниям-операторам социальных сетей осуществлять модерацию контента в соответствии со стандартами и правилами информационных платформ, в том числе посредством идентификации спама, использования технологий сличения входных и выходных данных хеш-кодов (например, использование цифровых отпечатков в целях выявления контента, имеющего отношение к террористической деятельности или эксплуатации детей), фильтрации сайтов по ключевым

¹⁰ Carly Nyst and Nick Monaco, *State-Sponsored Trolling: How Governments are Deploying Disinformation as Part of Broader Digital Harrassment Campaigns* (Palo Alto, Institute for the Future, 2018).

¹¹ World Wide Web Foundation, "Artificial intelligence: the road ahead in low- and middle-income countries", Washington, D.C., 2017.

¹² Zeynep Tufekci, "YouTube, the great radicaliser", *New York Times*, 10 March 2018. Имеется по адресу: <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>; James Williams, *Stand Out of our Light: Freedom and Resistance in the Attention Economy* (Cambridge, Cambridge University Press, 2018).

¹³ В последнее время ряд технологических платформ заявили о своем намерении отказаться от «удержания», повышая уровень персонализации и уделяя приоритетное внимание качеству пользовательского опыта в онлайн-режиме; см. Julia Carrie Wong, "Facebook overhauls News Feed in favour of 'meaningful social interactions'", *The Guardian*, 11 January 2018. URL: <https://www.theguardian.com/technology/2018/jan/11/facebook-news-feed-algorithm-overhaul-mark-zuckerberg>. Однако без обеспечения надлежащего уровня прозрачности, отчетности и достаточного объема количественных показателей, позволяющих продемонстрировать, как системы ИИ формируют соответствующие оценки и реализуют вытекающие из них решения, трудно сделать вывод относительно того, в какой степени внесение соответствующего изменения оказывает ощутимое воздействие на опыт работы пользователей в Интернете.

словам, обработки естественного языка (путем оценки характера контента с точки зрения использования запрещенных слов или изображений) и других алгоритмов идентификации. Искусственный интеллект может использоваться в целях доведения до сведения владельцев учетных записей предупреждений, уведомлений о приостановке или закрытии доступа к услугам на основании нарушения условий предоставления услуг или блокировки или фильтрации веб-сайтов, использующих запрещенных на соответствующих доменах данные или контент. Компании-операторы социальных сетей используют искусственный интеллект для фильтрации контента с учетом всего набора своих правил (предусматривающих в том числе запрет на изображение обнаженного тела, домогательства, использование ненавистнических высказываний и т. д.), хотя степень использования такими компаниями систем автоматизации без участия человека в конкретных случаях остается неизвестной¹⁴.

14. Представители как частного, так и государственного секторов активно выступают за повышение роли искусственного интеллекта. Компании утверждают, что объем незаконного, нежелательного и вредоносного контента в Интернете намного превышает возможности его модерации человеком и утверждают, что искусственный интеллект является инструментом, способным помочь в нахождении более эффективного решения этой проблемы. По мнению операторов ряда платформ, искусственный интеллект не только позволяет более эффективно идентифицировать нежелательный (согласно установленным правилам), незаконный и подлежащий удалению контент (обычно это делает человек-модератор), но и обеспечивает более высокую степень точности, по сравнению с решениями, принимаемыми с участием человека. Со своей стороны государства выступают за создание эффективной, оперативной и автоматизированной системы модерации для решения ряда конкретных проблем, в том числе проблемы сексуального насилия в отношении детей и идентификации контента, относящегося к террористической деятельности, где искусственный интеллект уже широко используется, а также проблем, относящихся к нарушению авторских прав и удалению контента экстремистского или ненавистнического содержания¹⁵. Рекомендация Европейской комиссии о мерах по дальнейшему повышению эффективности борьбы против незаконного содержания в Интернете от марта 2018 года призывает Интернет-платформы использовать автоматические фильтры в целях выявления и удаления контента, имеющего отношение к террористической деятельности, и использовать в некоторых случаях надзор со стороны человека для обеспечения необходимого контроля в целях предотвращения

¹⁴ Инструмент приложения «Инстаграм», именуемый «ДипТекст», используется для определения «токсичности» контекста, позволяет пользователям адаптировать используемые ими фильтры слов и эмодзи и использует анализ взаимоотношений между пользователями для более четкого установления контекста (например, для определения того, насколько высказанный комментарий является просто шуткой между друзьями). Andrew Hutchison, “Instagram’s rolling out new tools to remove ‘toxic comments’”, Social Media Today, 30 June 2017. URL: <https://www.socialmediatoday.com/social-networks/instagrams-rolling-out-new-tools-remove-toxic-comments>.

¹⁵ Согласно сообщению, Соединенное Королевство разработало инструмент для автоматического обнаружения и удаления террористического веб-контента на этапе загрузки. См. например, Home Office, “New technology revealed to help fight terrorist content online”, press release, 13 February 2018. См. European Commission, proposal for a directive of the European Parliament and of the Council on Copyright in the Digital Single Market, COM(2016) 593 final, art. 13; Letter from the Special Rapporteur to the President of the European Commission, reference No. OL OTH 41/2018, 13 June 2018. URL: <https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-OTH-41-2018.pdf>.

неизбежных ошибок, возникающих в результате функционирования автоматизированных систем¹⁶.

15. Усилия по автоматизации модерации контента могут создавать риски принятия мер в нарушение права человека (A/HRC/38/35, см. пункт 56) Возможности модерации контента с помощью искусственного интеллекта ограничиваются рядом факторов, в частности трудностями, возникающими при работе с контекстом, широким диапазоном языковых ключей и значений, а также с лингвистическими и культурными особенностями. В силу того, что в основе приложений, использующих искусственный интеллект, часто лежат массивы данных, содержащих дискриминационные предположения¹⁷, и в условиях низкого уровня опасности чрезмерной модерации, существует высокая степень риска того, что такие системы будут удалять онлайн-контент по умолчанию или временно отключать учетные записи, которые не являются проблемными, и что контент будет удаляться на основании необъективных или дискриминационных представлений. Таким образом, в результате использования систем модерации контента, основанных на искусственном интеллекте, в наиболее неблагоприятном положении окажутся представители уязвимых групп. Например, инструмент «ДипТекст» приложения «Инстаграм» определил слово «мексиканец» как оскорбление, поскольку его использование было связано с массивами данных, в которых «мексиканец» ассоциировался со словом «незаконный», что привело к включению этого слова в код алгоритма в негативной коннотации¹⁸.

16. Искусственный интеллект затрудняет проверку логики, лежащей в основе принятия решений относительно конкретного контента. Даже в тех случаях, когда алгоритмический контент дополняется функцией контроля со стороны человека – мерой, которая, как утверждают представители крупных платформ-операторов социальных сетей, все труднее осуществлять с учетом масштабов их деятельности – при принятии решений предпочтение, как правило, отдается решениям, принятым с помощью машин (с учетом предположений относительно объективности, о чем говорилось выше), что мешает анализу результатов модерации контента, особенно в тех случаях, когда технический дизайн системы не позволяет обеспечивать надлежащий уровень транспарентности.

17. **Профилирование, реклама и таргетирование.** Достижения и перспективы в области искусственного интеллекта связаны с дальнейшим развитием бизнес-модели обработки данных в Интернете, в рамках которой пользователи получают доступ к бесплатному контенту, оплачивая его предоставлением своих личных данных. Располагая обширными массивами данных, накопленными за годы онлайн-мониторинга и сбора информации, компании могут вводить в системы искусственного интеллекта обширные массивы данных для разработки все более точных прогнозов и моделей таргетирования. Сегодня

¹⁶ Commission recommendation of 1 March 2018 on measures to effectively tackle illegal content online (C(2018) 1177 final). URL: <https://ec.europa.eu/digital-single-market/en/news/commission-recommendation-measures-effectively-tackle-illegal-content-online>; см. также Daphne Keller, Daphne Keller, “Comment in response to European Commission’s March 2018 recommendation on measures to further improve the effectiveness of the fight against illegal content online”, Stanford Law School, Center for Internet and Society, 29 March 2018. URL: <http://cyberlaw.stanford.edu/publications/comment-response-european-commissions-march-2018-recommendation-measures-further>.

¹⁷ См. Aylin Caliskan, Joanna Bryson and Arvind Narayanan, “Semantics derived automatically from language corpora contain human-like biases”, *Science*, vol. 356, No. 6334 (14 April 2017); Solon Barocas and Andrew Selbst, “Big data’s disparate impact”, *California Law Review*, vol. 104, No. 671 (2016).

¹⁸ Nicholas Thompson, “Instagram’s Kevin Systrom wants to clean up the &#%@! Internet”, *Wired*, 14 August 2017. URL: <https://www.wired.com/2017/08/instagram-kevin-systrom-wants-to-clean-up-the-internet/>.

реклама частных и государственных рекламодателей может осуществляться на индивидуальном уровне; потребители и избиратели выступают в роли объектов «микротаргетирования» с целью учета и использования их индивидуальных особенностей.

18. Таргетирование на основе искусственного интеллекта стимулирует сбор и использование больших объемов личных данных и повышает риски распространения дезинформации в целях манипулирования отдельными пользователями. Таргетирование может способствовать сохранению дискриминации, а также лишению пользователей доступа к информации и возможностям, например, посредством предоставления целенаправленной рекламы работы и жилья, которая исключает пожилых работников, женщин и представителей этнических меньшинств¹⁹. В результате отдельные лица не получают доступ к сбалансированной и разнообразной информации, например, в политических сообщениях; напротив, использование в социальных сетях микротаргетирования позволяет сформировать мировосприятие, непригодное для плюралистической политической дискуссии.

III. Правовая основа в области прав человека для искусственного интеллекта

A. Объем обязательств в области прав человека в контексте искусственного интеллекта

19. Инструменты искусственного интеллекта, как и все технологии, должны проектироваться, разрабатываться и развертываться в соответствии с обязательствами государств и обязанностям частных субъектов согласно международному праву прав человека. Право прав человека налагает на государства как негативные обязательства, предусматривающие отказ от мер, которые препятствуют осуществлению права на свободу мнений и их свободное выражение, так и позитивные обязательства по реализации права на свободу мнений и их свободное выражение и обеспечение его осуществления.

20. Что касается частного сектора, то государства обязаны гарантировать уважение индивидуальных прав²⁰, в частности права на свободу мнений и их свободное выражение, в том числе путем защиты людей от нарушающих нормы актов, совершаемых частными сторонами (статья 2 (1) Международного пакта о гражданских и политических правах). Государства могут выполнить это обязательство путем принятия юридических мер в целях ограничения или воздействия на разработку и внедрение приложений искусственного интеллекта посредством осуществления политики закупок государственными субъектами приложений искусственного интеллекта у частных компаний, использования схем саморегулирования и совместного регулирования и укрепления потенциала компаний частного сектора в сфере признания и приоритизации права на свободу мнений и их свободное выражение в рамках своей корпоративной деятельности.

¹⁹ Julia Angwin, Noam Scheiber and Ariana Tobin, “Dozens of companies are using Facebook to exclude older workers from job ads”, ProPublica, 20 December 2017. Имеется по адресу: <https://www.propublica.org/article/facebook-ads-age-discrimination-targeting>; Julia Angwin, Ariana Tobin and Madeleine Varner, “Facebook (still) letting housing advertisers exclude users by race”, ProPublica, 21 November 2017. URL: <https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin>.

²⁰ См. Руководящие принципы предпринимательской деятельности в аспекте прав человека: осуществление рамок Организации Объединенных Наций, касающихся «защиты, соблюдения и средств правовой защиты» (A/HRC/17/31) и A/HRC/38/35, пункты 6–8).

21. Компании также несут обязательства согласно праву прав человека, которые они должны учитывать в своей деятельности по разработке, внедрению и использованию технологий искусственного интеллекта (A/HRC/38/35, пункт 10). Руководящие принципы предпринимательской деятельности в аспекте прав человека: осуществление рамок Организации Объединенных Наций, касающихся «защиты, соблюдения и средств правовой защиты» обеспечивают «всеобщую норму ожидаемого поведения всех предприятий, где бы они ни осуществляли свою деятельность» (принцип 11), включая социальные сети и поисковые компании. В целях адаптации выводов Руководящих принципов к области искусственного интеллекта (там же, пункт 11), Руководящие принципы содержат требование о том, чтобы компании, как минимум, на высоком политическом уровне брали обязательства уважать права человека пользователей применительно ко всем приложениям искусственного интеллекта (принцип 16); избегали мер, способных нанести или способствующих нанесению ущерба правам человека в рамках использования технологии искусственного интеллекта, и предотвращали и смягчали любые негативные последствия, связанные с их деятельностью (принцип 13); проявляли должную осмотрительность в отношении систем искусственного интеллекта для выявления и устранения фактических и потенциальных последствий для осуществления прав человека (принципы 17-19); участвовали в осуществлении стратегий предотвращения и уменьшения последствий (принцип 24); проводили постоянный обзор деятельности, связанной с искусственным интеллектом, в том числе в рамках проведения консультаций с заинтересованными сторонами и общественностью (принципы 20-21) и представляли доступные средства правовой защиты в целях устранения негативных последствий использования систем искусственного интеллекта для осуществления прав человека (принципы 22, 29 и 31).

В. Право на свободу мнений

22. Свобода беспрепятственно придерживаться своих убеждений является абсолютным правом, закрепленным в статье 19 (1) Международного пакта о гражданских и политических правах и в статье 19 Всеобщей декларации прав человека. Оно «не допускает каких-либо исключений или ограничений», «ни законом, ни другими формами власти»²¹. В моем докладе Совету по правам человека 2015 года по вопросу об использовании средств шифрования и анонимности при передаче цифровых сообщений (A/HRC/29/32) я отметил, что в эпоху цифровых технологий способы хранения, передачи и защиты информации оказывают уникальное воздействие на осуществление права придерживаться своих убеждений. Хранимые в облаке поисковые запросы, история навигации в Интернете, электронная почта и текстовые сообщения, а также документы и записи на память в совокупности представляют собой виды цифровой деятельности и документации, которые составляют суть убеждений пользователей (там же, пункт 12). Как государственные, так и негосударственные субъекты могут создавать помехи функционированию этих механизмов и осуществлению процессов формирования и сохранения убеждений.

23. «Свобода формировать мнения и развивать их путем умозаключений»²² является одним из важнейших элементов осуществления права на свое мнение. Комитет по правам человека пришел к выводу о том, что осуществление этого права не допускает неоправданного принуждения в процессе формирования

²¹ Комитет по правам человека, замечание общего порядка № 34 (2011) о свободе мнений и их выражения, пункт 9, URL: www2.ohchr.org/english/bodies/hrc/docs/GC34.pdf; Manfred Nowak, *U.N. Covenant on Civil and Political Rights: CCPR Commentary* (1993).

²² Nowak, *U.N. Covenant on Civil and Political Rights*.

убеждений, идеологии, реакций и позиций индивида²³. В этой связи, меры принудительного неврологического вмешательства, программы идеологической обработки (например, «лагеря перевоспитания») или угрозы насилия, имеющие целью заставить граждан сформировать или изменить конкретные убеждения, нарушают статью 19 (1) Пакта. Комитет также считает, что принудительное «предоставление стимулов в виде особого обращения» может представлять собой форму убеждения, которая нарушает право на формирование и сохранение убеждений (см. [CCPR/C/78/D/878/1999](#)).

24. Использование технологий и курирование контента порождает новые вопросы относительно видов принуждения или побуждения, которые могут рассматриваться как вмешательство в право на формирование собственного мнения. Курирование контента уже давно используется в целях оказания воздействия на формирование убеждений индивида: например, в средствах массовой информации на первой странице размещаются определенные сообщения с целью формирования индивидуальных представлений о важных новостях дня и оказания на них влияния. Коммерческая реклама также призвана стимулировать формирование положительных представлений о конкретных продуктах и услугах и желание их приобрести.

25. Использование искусственного интеллекта расширяет традиционные возможности курирования контента в Интернете, обеспечивая более современные и эффективные средства персонализации и курирования контента для пользователей, которые своим охватом значительно превосходят возможности традиционных средств массовой информации. Преобладание конкретных методов курирования контента с помощью искусственного интеллекта вызывает обеспокоенность по поводу того, как его использование влияет на способность отдельных лиц формировать и сохранять свои убеждения. Например, несколько технологических компаний заявляют о том, что они выполняют подавляющее большинство поисковых запросов в режиме онлайн. В результате существования корпоративной монополии на рынке поисковых систем пользователям крайне трудно отказаться от алгоритмического ранжирования и курирования результатов поиска и (как того и желают компании) у них может сформироваться представление о том, что полученные результаты являются наиболее значимой или объективной информацией по конкретному вопросу. Отсутствие прозрачности в отношении принципов разработки и использования критериев поиска на основе использования искусственного интеллекта может также укрепить предположение о том, что результаты поиска, полученные на определенной платформе, представляют собой объективное представление фактической информации.

26. Поэтому вопросы, которые возникают в сфере использования искусственного интеллекта в целях курирования контента в связи с доминирующим положением на рынке, ставят под сомнения сложившиеся в исторической перспективе представления о том, каким образом курирование контента оказывает или не оказывает воздействия на способность формировать собственные убеждения. Новизна поднятых вопросов в сочетании с общим отсутствием правовых норм, касающихся ограничения права на свободу мнений, порождает больше вопросов, чем ответов в отношении влияния курирования контента при помощи искусственного интеллекта на осуществление прав человека в современной цифровой среде. Тем не менее эти вопросы должны стимулировать исследования, ориентированные на изучение социальных, экономических и политических последствий для осуществления прав человека использования курирования контента при помощи искусственного интеллекта. В рамках проектирования и

²³ *Ён Ю Кан против Республики Корея*, сообщение Совета по правам человека No. 878/1999, 16 июля 2003 года ([CCPR/C/78/D/878/1999](#)).

разработки соответствующих систем искусственного интеллекта компании должны, как минимум, предоставлять конкретную информацию о том, как они разрабатывают и используют на своих платформах критерии курирования и персонализации контента, в том числе о стратегии и процедурах выявления социальных, культурных или политических предубеждений.

С. Право на свободу выражения мнений

27. Статья 19 (2) Пакта гарантирует расширительное право на «поиск, получение и распространение всякого рода информации и идей», которое необходимо защищать и соблюдать независимо от государственных границ и любыми средствами. Осуществление права на свободу выражения мнений в конечном итоге тесно связано с осуществлением других прав и служит основой для эффективного функционирования демократических институтов, и, следовательно, защита, соблюдение и поощрение права на свободу выражения мнений влечет за собой обязанность включать поощрение многообразия и независимости средств массовой информации и защиты доступа к информации²⁴.

28. В отличие от права на формирование и сохранение убеждений, права на выражение мнений и доступ к информации и идеям могут в некоторых случаях ограничиваться (статья 19 (3) Пакта). Ограничения должны соответствовать нормам законности, что означает, что они предусматриваются в явной форме в законе, который удовлетворяет требованиям ясности и четкости и подлежит толкованию со стороны независимых судебных органов; необходимости и соразмерности, что означает, что они являются наименее ограничительной мерой, необходимой для обеспечения конкретного законного интереса и не ставят под угрозу суть права; и легитимности, что означает, что они должны защищать упомянутый законный интерес, а именно права и репутацию других лиц, государственную безопасность или общественный порядок, или здоровье и нравственность населения (A/HRC/38/35, пункт 7). В этом контексте выражение мнений может быть ограничено на основании статьи 20 (2) Пакта, которая требует от государств запрета «всякого выступления в пользу национальной, расовой или религиозной ненависти, представляющего собой подстрекательство к дискриминации, вражде или насилию»; однако, ограничения также должны удовлетворять совокупным критериям законности, необходимости и легитимности²⁵.

29. Сложность принятия решений, связанных с модерацией контента, может еще больше усугубляться внедрением автоматизированных процессов. В отличие от людей, сегодня алгоритмы не в состоянии дать оценку культурному контексту, идентифицировать иронию или осуществить критический анализ, необходимый для точного определения, например, контента «экстремистского характера» или ненавистнические высказывания и, в результате этого, они по умолчанию скорее всего обеспечат включение механизма блокировки и ограничения

²⁴ Специальный докладчик по вопросу о поощрении и защите права на свободу мнений и их свободное выражение, Организации по безопасности и сотрудничеству в Европе по вопросу о свободе средств массовой информации, Специальный докладчик Организации американских государств по вопросу о свободе выражения мнений и Специальный докладчик Африканской комиссии по правам человека и народов по вопросу о свободе выражения мнений и доступе к информации, «Совместная декларация о свободе выражения мнений, «фейковых новостях», дезинформации и пропаганде», 3 марта 2017 года. URL: <https://www.osce.org/fom/302796>; см. также Комитет по правам человека, замечание общего порядка № 34 (2011) о свободе мнений и их выражения; A/HRC/29/32, пункт 61, и A/HRC/32/38, пункт 86.

²⁵ Комитет по правам человека, замечание общего порядка № 34 (2011) о свободе мнений и их выражения, пункт 50.

контента, ограничивая тем самым право отдельных пользователей на выражение собственного мнения, а также их права на доступ к информации без каких-либо ограничений и цензуры²⁶.

30. В системе, управляемой с помощью искусственного интеллекта, распространение информации и идей регулируется непрозрачными механизмами, использующими приоритеты, которые могут противоречить задаче создания благоприятной среды для поощрения разнообразия средств массовой информации и независимых мнений. С учетом этого Комитет по правам человека пришел к выводу о том, что государства-участники должны принять все необходимые меры ... для того, чтобы предупредить случаи монополизации, ненадлежащее доминирование в СМИ частных медиа групп или сосредоточение СМИ в их руках, что может негативно сказаться на разнообразии источников информации и взглядов²⁷.

31. Когда речь идет о платформах и веб-сайтах, использующих системы искусственного интеллекта, пользователи также не имеют доступа к правилам игры. Отсутствие ясности в отношении масштабов применения искусственного интеллекта и алгоритмических приложений не позволяет понять, когда и по каким метрическим критериям происходит распространение, ограничение или таргетирование информации. Небольшие уступки в целях решения этой проблемы, такие, как избирательная идентификация спонсированных результатов поиска²⁸ или помещение в социальных сетях информации о размещении рекламы, оплаченной политическими деятелями, могут отчасти помочь пользователям понять правила функционирования информационной среды, однако они не позволяют отразить или устранить вопросы, которые возникают в отношении масштабов использования алгоритмических процессов, определяющих эту среду.

32. Даже в тех случаях, когда отдельные лица получают информацию о наличии, сфере охвата и функционировании системы искусственного интеллекта, эти системы могут свести на нет усилия по обеспечению транспарентности и адекватности. На сегодняшний день никаких современных и масштабируемых средств для изучения и транспарентного представления технических принципов принятия автоматизированных решений в онлайн-сфере разработано не было²⁹. Это означает, что нередко права людей на выражение мнений ущемляются без предоставления им возможности изучить или понять, почему, каким образом и на каком основании это происходит.

D. Право на неприкосновенность частной жизни

33. Право на неприкосновенность частной жизни нередко выступает в качестве залога свободы мнений и их свободного выражения³⁰. Статья 17 Пакта защищает индивидуума от «произвольного или незаконного вмешательства в его или ее личную и семейную жизнь, произвольных или незаконных посягательств на неприкосновенность жилища или тайну корреспонденции» и «незаконных

²⁶ Council of Europe, *Algorithms and Human Rights*.

²⁷ Комитет по правам человека, замечание общего порядка № 34 (2011) о свободе мнений и их выражения, пункт 40.

²⁸ Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York, New York University Press, 2018).

²⁹ Mike Ananny and Kate Crawford, "Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability", *New Media and Society*, vol. 20, No. 3 (13 December 2016). URL: <http://journals.sagepub.com/doi/abs/10.1177/1461444816676645?journalCode=nmsaa>.

³⁰ См. A/HRC/29/32, см. пункт 16, резолюция 68/167 Генеральной Ассамблеи и резолюция 20/8 Совета по правам человека.

посягательств на его честь и репутацию» и предусматривает, что «каждый человек имеет право на защиту закона от такого вмешательства или таких посягательств». Управление Верховного комиссара по правам человека и Совет по правам человека подчеркнули, что при любом вмешательстве в осуществление права на неприкосновенность частной жизни необходимо соблюдать принципы законности, необходимости и соразмерности (A/HRC/27/37, пункт 23 и резолюция 34/7 Совета по правам человека, пункт 2).

34. Эффективность систем принятия решений, основанных на использовании искусственного интеллекта, зависит от сбора и обработки данных, от контекстуальных, обезличенных данных до персональной идентификационной информации, причем используемые основные массивы данных, поступающих в системы искусственного интеллекта, находятся где-то посередине, данных, логически выводимых или извлеченных из личных данных или личные данные, подвергнутые анонимизации (часто недостаточно совершенным образом). Компании используют данные, полученные в результате онлайн-профилирования и использования цифровых идентификаторов, закупают наборы данных у третьих сторон, например, у брокеров данных, и получают новые данные из больших объемов агрегированных данных, предназначенных для ввода в системы искусственного интеллекта. Потребительские товары и автономные системы, использующие элементы искусственного интеллекта, нередко оснащены датчиками, которые генерируют и собирают обширные объемы данных о лицах в непосредственной близости от них³¹, а в социальных сетях используются методы искусственного интеллекта в целях получения посредством анализа конфиденциальной информации о лицах, которые эту информацию не представляли и не подтверждали, в том числе о сексуальной ориентации, семейных отношений, религиозных убеждениях, состоянии здоровья или политической принадлежности.

35. Искусственный интеллект разрушает традиционные представления о таких понятиях, как согласие, цели и применение ограничений, транспарентность и подотчетность, которые являются основой для международных стандартов защиты данных³². Поскольку системы искусственного интеллекта построены на использовании существующих и создания новых наборов данных, в контексте искусственного интеллекта частным лицам в практическом плане не имеет смысла знать, понимать и контролировать, как используются их данные. После перепрофилирования данных в системе искусственного интеллекта, они утрачивают свой первоначальный контекст, что повышает риск того, что данные о лицах станут неточными или устаревшими, и лишает лиц возможности вносить в данные исправления или удалять их. Системы, функционирующие на основе искусственного интеллекта, используются для принятия вытекающих из этих данных решений, которые могут оказывать большое влияние на жизнь отдельных лиц³³, и, тем не менее, частные лица располагают ограниченными возможностями контроля информации, полученной из их личных данных, в силу неэффективности существующей технологии анонимизации данных.

Е. Обязательства, касающиеся недискриминации

36. Недискриминация является неотъемлемым принципом в области права прав человека, который не только подтверждает существование у государств обязательств обеспечивать осуществление всех других прав человека без какой-

³¹ Article 19 and Privacy International, "Privacy and freedom of expression".

³² См. Комитет по правам человека, замечание общего порядка № 16 (1988) о праве на неприкосновенность частной жизни, пункт 10.

³³ Article 19 and Privacy International, "Privacy and freedom of expression".

либо дискриминации, но и, как это предусмотрено в статье 26 Пакта, является самостоятельной гарантией равенства перед законом и равной защиты закона. В соответствии с прямой обязанностью государств, «всякого рода дискриминация должна быть запрещена законом и закон должен гарантировать всем лицам равную и эффективную защиту против дискриминации по какому бы то ни было признаку, как-то: расы, цвета кожи, пола, языка, религии, политических или иных убеждений, национального или социального происхождения, имущественного положения, рождения или иного обстоятельства». Таким образом, статьи 17 и 19 включают в индивидуальные права на свободу от дискриминации право на формирование и сохранение убеждений, представление идей и информации и доступа к ним и осуществление права на неприкосновенность частной жизни и защиту личных данных.

37. Искусственный интеллект может способствовать формированию и закреплению предвзятости и дискриминации и этот потенциал дискриминации сохраняется и в области осуществления права на свободу мнений и их свободное выражение. Алгоритмы модерации могут не обеспечить учет содержащихся в контенте культурных, языковых или гендерных контекстов и особенностей или государственных интересов³⁴. В то время как новостные ленты, функционирующие на основе алгоритмов искусственного интеллекта, могут способствовать закреплению дискриминационных представлений, системы искусственного интеллекта, используемые в целях профилирования и рекламы, продемонстрировали свою эффективность в плане дискриминации по расовому, религиозному и гендерному признаку³⁵. Функция искусственного интеллекта «Автоматическое заполнение» также приводит к получению дискриминационных в расовом отношении результатов³⁶.

38. Ряд факторов позволяет понять, как предвзятость встраивается в системы искусственного интеллекта, укрепляя их дискриминационный потенциал. В их число, в частности, входят решения, которые принимались при разработке систем искусственного интеллекта относительно источников происхождения и выборки наборов данных, используемых для обучения этих систем, социальные и культурные предубеждения, которые разработчики могут встраивать в эти наборы данных, сами модели искусственного интеллекта и конкретные способы представления результатов функционирования моделей искусственного интеллекта. Например, получение до 20 процентов ошибок при распознавании лиц женщин и людей с более темным цветом кожи объясняется тем фактом, что программы распознавания лиц используют баз данных, содержащих информацию в основном о белых мужчинах³⁷. В тех случаях, когда такие системы

³⁴ В частности, это привело к удалению исторических фотографий, обладающих особым культурным значением. См. Julia Carrie Wong, “Mark Zuckerberg accused of abusing power after Facebook deletes ‘napalm girl’ post”, *The Guardian*, 9 September 2016. URL: <https://www.theguardian.com/technology/2016/sep/08/facebook-mark-zuckerberg-napalm-girl-photo-vietnam-war>; см. также A/HRC/38/35, пункт 29.

³⁵ Julia Angwin, Madeleine Varner and Ariana Tobin, “Facebook enabled advertisers to reach ‘Jew haters’”, *ProPublica*, 14 September 2017. URL: <https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters>; Ariana Tobin, “Why we had to buy racist, sexist, xenophobic, ableist and otherwise awful Facebook ads,” *ProPublica*, 27 November 2017. URL: <https://www.propublica.org/article/why-we-had-to-buy-racist-sexist-xenophobic-ableist-and-otherwise-awful-facebook-ads>.

³⁶ Paris Martineau, “YouTube’s search suggests racist autocompletes”, *The Outline*, 13 May 2018. URL: <https://theoutline.com/post/4536/youtube-s-search-autofill-suggests-racist-results?zd=1&zi=3ygz6hw>.

³⁷ Joy Buolamwini, “The dangers of supremely white data and the coded gaze”, presented at Wikimania 2018, Cape Town. URL: <https://www.youtube.com/watch?v=ZSJXKoD6mA8&feature=youtu.be>.

используются, например, в целях классификации изображений с помощью поисковой системы, их дискриминационный потенциал может привести к реальному ограничению прав граждан на поиск, получение и распространение информации, свободу собраний и ассоциаций.

Е. Право на эффективное восстановление в правах

39. Международное право в области прав человека гарантирует лицам право на правовую защиту, которое устанавливается компетентными судебными, административными или законодательными властями (статья 2 (3) Пакта). Средства правовой защиты должны быть известны и доступны всем, кто утверждает, что их права были нарушены; включать в себя тщательное оперативное и беспристрастное расследование предполагаемых нарушений³⁸ и быть способны остановить длящиеся нарушения (A/HRC/27/37, пункты 39–41).

40. Системы искусственного интеллекта часто препятствуют осуществлению права на эффективное восстановление в правах. Во-первых, по своей сути они в принципе не предоставляют уведомления отдельным лицам. Практически во всех случаях применения технологий искусственного интеллекта в информационной среде люди не отдают себе отчет о масштабах, сфере охвата или даже существовании алгоритмических процессов принятия решений, которые могут оказывать влияние на осуществление их прав на свободу мнений и их свободное выражение. Второй и более сложный аспект связан с возможностью изучения самой системы искусственного интеллекта. Логика, лежащая в основе алгоритмического решения, может быть непонятна даже экспертам, знакомым с механизмом работы этой системы. Хотя логично предположить, что повышение уровня транспарентности системы искусственного интеллекта расширит возможности для ее изучения, транспарентность алгоритмов не обязательно обеспечивает вразумительные ответы на вопросы относительно процессов принятия решений. Алгоритмы могут скрывать механизм принятия последующего решения или быть настолько сложными и зависеть от контекста, что этот фактор будет препятствовать получению объяснения. Ситуация еще более усложняется тем, что действующие в информационной среде компании часто обновляют свои алгоритмы³⁹; аналогичным образом, программы машинного обучения с течением времени могут менять свои собственные правила и алгоритмы.

41. Эти проблемы усугубляются переходом к автоматизации самих систем правовой защиты, когда решения по жалобам отдельных пользователей на модерацию контента или негативные последствия для прав человека, обусловленные использованием технологии искусственного интеллекта, рассматриваются и выносятся при помощи технологий искусственного интеллекта⁴⁰. Автоматизированные процессы предоставления ответов заставляют сомневаться в эффективности механизмов рассмотрения жалоб в качестве средства правовой защиты, учитывая отсутствие дискреционных полномочий, контекстуального анализа и возможности принятия независимых решений, встроенных в такие процессы⁴¹.

³⁸ Комитет по правам человека, замечание общего характера № 31 (2004) о сути общего юридического обязательства, налагаемого на государства — участники Пакта, пункт 15.

³⁹ Barry Schwartz, “Google: we make thousands of updates to search algorithms each year”, Search Engine Roundtable, 5 June 2015. URL: <https://www.seroundtable.com/google-updates-thousands-20403.html>.

⁴⁰ Council of Europe, *Algorithms and Human Rights*.

⁴¹ Pei Zhang, Sophie Stalla-Bourdillon and Lester Gilbert, “A content-linking-context model for ‘notice-and-take-down’ procedures”, *WebSci '16*, May 2016. URL: <http://takedownproject.org/wp-content/uploads/2016/04/ContentLinkingModelZhangStallaGilbert.pdf>.

G. Законодательные, нормативные и политические меры реагирования на появление искусственного интеллекта

42. В настоящее время многие государства разрабатывают национальные программы в области искусственного интеллекта, с тем чтобы изучить и разработать стратегии и инициативы, призванные обеспечить гражданам максимальные выгоды от использования искусственного интеллекта⁴². Хотя ни одно государство еще не разработало всеобъемлющий закон или нормативный акт в целях регулирования искусственного интеллекта, в процессе разработки такого решения следует проявлять осмотрительность, поскольку оно может оказаться мало-пригодным для этой инновационной области и привести к тому, что в виду отсутствия детальной информации будут приняты чрезмерно ограничительные или слишком либеральные положения. Возможно целесообразнее использовать секторальное регулирование, хотя, можно утверждать, что существующие законы и нормативные акты, например в области защиты данных, обладают достаточной гибкостью и не требуют принятия дополнительных законодательных актов.

43. В то же время государства должны обеспечивать разработку искусственного интеллекта в соответствии со стандартами в области прав человека. Любые усилия, направленные на разработку государственной политики и нормативные положения в области искусственного интеллекта, должны обеспечить рассмотрение проблем, вызывающих беспокойство с точки зрения прав человека⁴³. В частности, право на свободу мнений и их свободное выражение часто не обсуждается в общественных дискуссиях и политических дебатах по вопросам искусственного интеллекта, в ходе которых при рассмотрении вопросов в области прав человека основное внимание, как правило, уделяется вопросам предвзятости и дискриминации в процессе предоставления услуг.

44. Поскольку разработка эффективных систем искусственного интеллекта зависит от приобретения больших массивов данных, а также долгосрочных инвестиций в создание технического потенциала, существует значительная вероятность того, что господствующие позиции в сфере разработки, производства и потенциала будут занимать структуры частного сектора, что приведет к росту зависимости компаний государственного сектора от таких компаний с точки зрения получения доступа к системам искусственного интеллекта. Такая тенденция скорее всего приведет к более тесному слиянию корпоративных и государственных интересов и формированию в этой связи прочных убеждений среди общественности. Такая ситуация особенно характерна для сегментов информационного пространства, таких как социальные медиа-платформы, поисковые системы и других технологии, где правительства часто выступают в роли пользователей, а не провайдеров услуг. Само по себе согласование общественных и частных интересов не ведет к ограничению прав человека, однако вызывает беспокойство в отношении прозрачности и подотчетности. По мере развития искусственного интеллекта в частном секторе возникает весьма реальный риск того, что государства будут делегировать частным компаниям решение все

⁴² Tim Dutton, "An overview of national AI strategies", Medium, 28 June 2018. URL: <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd..>

⁴³ Так, например, вызывает беспокойство тот факт, что парламентский комитет Соединенное Королевство Великобритании и Северной Ирландии опубликовал 200-страничный доклад, в котором нет ни одного упоминания прав человека. См. United Kingdom of Great Britain and Northern Ireland House of Lords Select Committee on Artificial Intelligence, "AI in the United Kingdom".

более сложных и серьезных задач, связанных с осуществлением цензуры и контроля.

45. В рамках своих попыток по созданию законодательной базы или разработке стратегии в области искусственного интеллекта любое государство должно принимать во внимание приложения искусственного интеллекта как в государственном, так и в частном секторе, а не заниматься регулированием искусственного интеллекта только в государственном секторе. В этой связи Совет Европы пришел к выводу о том, что «решение вопросов, касающихся управления с помощью алгоритмов, и/или разработка нормативных положений является прерогативой государственной политики и не должно отдаваться в распоряжение только частных субъектов»⁴⁴. В целях преодоления озабоченностей, возникающих в связи с использованием искусственного интеллекта, государства в рамках своей деятельности могут предъявлять компаниям повышенные требования в отношении транспарентности и раскрытия информации и разрабатывать эффективные законодательные акты о защите данных.

46. В государственном и частном секторах все большее распространение получают инициативы, направленные на изучение и интеграцию этических соображений в деятельность по приобретению, проектированию, внедрению и применению систем искусственного интеллекта. Я настоятельно рекомендую включить в эти усилия рассмотрение проблем в области прав человека. Нередко внимание, которое частный сектор уделяет этическим соображениям, и требования относительно соблюдения этических норм, которые выдвигает государственный сектор, нередко предполагает существование сопротивления принятию нормативных положений в области прав человека⁴⁵. В то время как этика является важнейшей основой для решения конкретных проблем в области искусственного интеллекта, она не подменяет права человека, которые каждое государство по закону обязано соблюдать. Компании и правительства должны обеспечить прочную интеграцию соображений и обязательств, касающиеся прав человека, во все аспекты функционирования искусственного интеллекта уже на данном этапе разработки этических кодексов и руководящих указаний⁴⁶.

IV. Правовая основа в области прав человека для искусственного интеллекта

47. В последних докладах мандатарий изложил компаниям соображения юридического и практического характера с целью включения правозащитных принципов в качестве центрального элемента их стратегий в области политики регулирования контента и представил информацию о материально-правовых стандартах и процессах, обеспечивающих выполнение компаниями своих обязательств в области прав человека во всех сферах своей деятельности в соответствии с Руководящими принципами предпринимательской деятельности в аспекте прав человека. Эти же принципы служат основой для предлагаемого здесь подхода в отношении технологий искусственного интеллекта. Описываемые ниже материально-правовые стандарты и процессы распространяются на компании, которые в процессе разработки, внедрения и использования систем искусственного интеллекта выступают в качестве субъектов, и на государства, которые обязаны воздерживаться от вмешательства в осуществление прав

⁴⁴ Council of Europe, *Algorithms and Human Rights*.

⁴⁵ Ben Wagner, "Ethics as an escape from regulation: from ethics-washing to ethics-shopping?", in *Being Profiling: Cogitas Ergo Sum*, Mireille Hildebrandt, ed. (Amsterdam University Press (forthcoming)).

⁴⁶ Article 19 and Privacy International, "Privacy and freedom of expression".

человека в ходе внедрения и использовании систем искусственного интеллекта. Данные стандарты и процедуры призваны обеспечить соблюдение правовых норм в области прав человека в качестве первоочередной задачи в сфере разработки искусственного интеллекта. В основе предлагаемых стандартов и процессов лежат два принципиальных требования: обеспечение защиты и уважения лица и его автономии, что является ключевым условием для осуществления права на свободу мнений и их свободное выражение; и важность раскрытия конкретных данных представителями государственного и частного секторов в рамках открытых и новаторских усилий в целях разъяснения обществу технологий искусственного интеллекта и оказания содействия их изучению.

A. Основные стандарты для систем искусственного интеллекта

48. В процессе разработки своих стандартов, правил и систем компании должны руководствоваться универсальными принципами в области прав человека (A/HRC/38/35, пункты 41–43). Адресованные обществу термины и руководящие указания следует дополнять обязательствами, которые предусматриваются внутренней стратегией, в целях учета правозащитных аспектов в рамках всей деятельности компаний, особенно в связи с разработкой и развертыванием программ искусственного интеллекта и алгоритмических систем. Компаниям следует рассмотреть вопрос о разработке профессиональных стандартов для инженеров, занимающихся разработкой искусственного интеллекта, в целях включения правозащитных обязательств в руководства по техническому проектированию и эксплуатации. Разработка кодексов этики и соответствующих институциональных структур может стать важным дополнением, но не заменой обязательств в области прав человека. В кодексах и руководящих указаниях, подготовленных органами как государственного, так и частного сектора, следует особо отметить, что право прав человека содержит основополагающие нормы, касающиеся защиты отдельных лиц в сфере искусственного интеллекта, в то время как этические рамки должны помочь в определении конкретного содержания и обеспечения прав человека в конкретных условиях.

49. Компании и правительства должны предоставить лицам четкие разъяснения относительно того, какие решения в информационной среде принимают автоматизированные системы, какие сопровождаются контролем со стороны человека и представить информацию относительно общих элементов логики, используемых этими системами. Отдельные лица должны также информироваться о том, что предоставляемые ими личные данные субъекту в частном секторе станут частью набора данных, используемых системой искусственного интеллекта (или непосредственно, или через услуги или через обращения к веб-сайту), с тем чтобы они могли принять эту информацию к сведению при принятии решения о предоставлении согласия на сбор данных и типах данных, которые они желают раскрывать⁴⁷. Подобно публичным уведомлениям, которые необходимо размещать при использовании камер видеонаблюдения, системы искусственного интеллекта должны в четкой и понятной форме активно информировать пользователей (с помощью инновационных средств, таких как всплывающие сообщения), что они участвуют в процессе принятия решений с использованием искусственного интеллекта или предоставляют данные для этого процесса, а также сообщить им конкретную информацию о логике этого процесса и значении ее последствий для данного лица.

⁴⁷ См. Комитет по правам человека, замечание общего порядка № 16 (1988) о праве на неприкосновенность частной жизни.

50. Транспарентность не ограничивается предоставлением отдельным пользователям информации о наличии технологий искусственного интеллекта на используемых ими платформах и в онлайн-сервисах. Компаниям и правительствам необходимо обеспечить транспарентность по всей цепочке функционирования искусственного интеллекта. Эффективное обеспечение прозрачности не обязательно должно быть сопряжено с решением сложных задач; даже простое разъяснение целей, стратегий, входных и выходных параметров искусственного интеллекта может способствовать просвещению общественности и развитию публичного диалога⁴⁸. Вместо того, чтобы решать сложную задачу популяризации сложных технических процессов в аудитории неспециалистов, компании должны стремиться достичь цели обеспечения транспарентности путем предоставления нетехнических сведений о функционировании системы. В этой связи особое внимание следует уделять информированию отдельных пользователей о существовании, предназначении, устройстве и воздействии системы искусственного интеллекта, а не сообщать информацию об исходном коде, калибровочных данных и данных на входе и выходе системы⁴⁹.

51. Обеспечение полной транспарентности в отношении последствий использования системы искусственного интеллекта в информационной среде требует раскрытия информации, например, об объеме данных, удаляемых системой искусственного интеллекта, числе случаев утверждения модератором предложенных искусственным интеллектом решений об удалении контента, количестве решений об удалении контента, которые становятся предметом споров, и случаев успешно оспоренных решений. Пользователи должны иметь доступ к агрегированным данным, иллюстрирующим тенденции в сфере представления контента, а также к тематическим исследованиям, позволяющим понять логику приоритизации определенного контента по сравнению с другим. Важнейшим элементом полной транспарентности является раскрытие информации об источниках финансирования и бенефициарах политической и коммерческой рекламы. Субъекты государственного и частного секторов, использующие системы искусственного интеллекта, должны также раскрывать информацию о предельных возможностях системы искусственного интеллекта, в том числе, о мерах доверия, известных сценариях возникновения нештатных ситуаций и соответствующих ограничениях на использование⁵⁰.

52. Борьба с распространением дискриминации в системах искусственного интеллекта представляет собой жизненно важную задачу для компаний и правительств; неспособность решить проблему удаления дискриминационных элементов и негативного воздействия сделает эти технологии не только неэффективными, но и опасными. При рассмотрении путей устранения предвзятости и дискриминации в системах искусственного интеллекта компании и правительства могут опираться на значительных накопленный интеллектуальный и ресурсный потенциал; в целом речь идет о необходимости идентификации и учета элементов дискриминации как на входе, так и выходе системы. Решение этой задачи предполагает, как минимум, устранение ошибок в выборке (когда массивы данных некорректно отражают ситуацию в обществе), проведение очистки данных в целях устранения дискриминационного контента и принятия мер в целях компенсации данных, которые «содержат характеристики

⁴⁸ Aaron Rieke, Miranda Bogen and David Robinson, “Public scrutiny of automated decisions: early lessons and emerging methods” (Omidyar and Upturn, 2018).

⁴⁹ Rieke, Bogen and Robinson, “Public scrutiny of automated decisions”.

⁵⁰ Amnesty International and Access Now, “Toronto declaration: protecting the right to equality and non-discrimination in machine learning systems”, art. 27 (d), 2018. URL: <https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/>.

исторических и структурных моделей дискриминации»⁵¹ и весьма вероятно позволят системам искусственного интеллекта построить дискриминирующие прокси-алгоритмы. Активный мониторинг дискриминирующих результатов функционирования систем искусственного интеллекта также является элементом предотвращения и смягчения неблагоприятных последствий для прав человека отдельных лиц.

В. Процессы для систем искусственного интеллекта

53. Оценки воздействия на осуществление прав человека. В целях обеспечения максимальной прозрачности систем искусственного интеллекта компании и правительства должны применять меры по обеспечению доступа к таким системам в целях осуществления анализа и контроля на всех этапах жизненного цикла от разработки концепции до внедрения. Одним из способов обеспечения прозрачности является подготовка оценок воздействия систем искусственного интеллекта с точки зрения их готовности к решению правозащитных задач, и эта работа должна проводиться до начала закупочной деятельности, перед этапом разработки и внедрения, и предусматривать проведение самооценки и внешней аттестации. Аналитический центр AI Now («ИИ прямо сейчас») выступил с предложением о проведении оценки алгоритмического воздействия государственным учреждением, в рамках которой правительства готовят внутренний обзор систем искусственного интеллекта и в целях проверки гипотез и выводов оказывают содействие в проведении внешней оценки⁵². Компании должны также проводить аналогичные оценки.

54. В государственном секторе закупки технологий искусственного интеллекта у частных поставщиков должны проводиться после проведения консультаций с общественностью в целях выяснения мнений и предложений относительно разработки и внедрения системы искусственного интеллекта. Компании и правительства должны постоянно проводить конструктивные консультации с представителями гражданского общества, правозащитных групп, соответствующих местных общин и исторически маргинализованных и недопредставленных групп населения перед разработкой, приобретением и началом использования систем и технологий искусственного интеллекта.

55. Аудиторские проверки. Проведение внешнего обзора систем искусственного интеллекта является надежной гарантией точности, независимости и прозрачности оценки. Поэтому проведение регулярных независимых аудиторских проверок закупочной деятельности должно дополнять подготовку предварительных оценок воздействия на осуществление прав человека в качестве одного из важных механизмов обеспечения прозрачности и подотчетности системы искусственного интеллекта. Представители частного сектора ставят под сомнение возможность проведения аудиторских проверок в сфере искусственного интеллекта, ссылаясь на необходимость защиты патентованных технологий. В то время как эти сомнения могут быть вполне обоснованы, я разделяю мнение «ИИ прямо сейчас» о том, что отказ поставщика обеспечивать прозрачность в рамках функционирования системы противоречит собственным обязательствам государственного органа в отношении подотчетности, особенно в тех случаях, когда приложение искусственного интеллекта используется учреждением государственного сектора.

⁵¹ Iason Gabriel, “The case for fairer algorithms”, Medium, 14 March 2018. URL: https://medium.com/@Ethics_Society/the-case-for-fairer-algorithms-c008a12126f8.

⁵² Dillon Reisman and others, “Algorithmic impact assessments: a practical framework for public agency accountability” (AI Now, 2018). URL: <https://ainowinstitute.org/aiareport2018.pdf>.

56. В любом случае существует ряд новаторский предложений относительно проведения аудиторских проверок технологий искусственного интеллекта, позволяющих защитить права интеллектуальной собственности; так, например, алгоритмический протокол может генерировать доказательства с нулевым разглашением информации в целях подтверждения их соответствия определенным параметрам, что устраняет необходимость проведения тщательного изучения лежащего в основе операции алгоритма⁵³; или алгоритмические данные могут быть раскрыты экспертам третьих сторон, которые берут их на хранение при условии соблюдения конфиденциальности, что позволит установить соответствие алгоритма задаче защиты общественных интересов и одновременно не допустить его попадания в открытый доступ⁵⁴. В конфиденциальном порядке доступ к системам искусственного интеллекта может быть предоставлен государственным органам, занимающимся регулированием сектора связи и защитой конкуренции, как это уже происходит, например, в сфере регулирования деятельности игорных заведений в Австралия и Новой Зеландии, где в рамках проведения аудиторских проверок компании должны регулярно предоставлять доступ регулирующим органам к алгоритмическим системам⁵⁵. В научной литературе содержатся и другие предложения относительно новаторских форм проведения аудиторский проверок искусственного интеллекта⁵⁶.

57. При использовании каждого из этих механизмов могут возникать проблемы, особенно в информационной среде, однако компании должны прилагать все усилия в целях создания условий для проведения аудиторских проверок систем искусственного интеллекта. В целях повышения эффективности проверок правительства должны принимать политические решения и законы, которые будут заставлять компании разрабатывать аудируемый код искусственного интеллекта, гарантировать сохранение истории проведения аудиторских проверок и тем самым обеспечивать затронутым лицам возможность получения доступа к более транспарентной информации.

58. **Личная автономия.** Способность отдельных лиц формировать и сохранять свои убеждения, получать доступ к идеям и выражать их в информационной среде не должна скрытым образом подменяться, подвергаться манипулированию или ограничиваться искусственным интеллектом. Уважение личной автономии, как минимум, означает, что пользователи обладают знаниями, имеют выбор и осуществляют контроль. Широко распространенные скрытые приложения искусственного интеллекта, использующие непрозрачные процессы представления контента, персонализации, модерации, профилирования и таргетирования, ограничивают возможности людей в сфере осуществления права на свободу мнений, их выражения и права на неприкосновенность частной жизни. Компании должны учитывать негативное воздействие на права человека приложений искусственного интеллекта, которые уделяют первоочередное внимание продвижению коммерческих или политических интересов, а не защите транспарентности и индивидуального выбора.

59. **Информирование и согласие.** Компании должны обеспечивать, чтобы пользователи были в полной мере информированы о том, как алгоритмы принятия решений влияют на то, как они используют платформы, веб-сайты или

⁵³ Council of Europe, *Algorithms and Human Rights*.

⁵⁴ Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, Harvard University Press, 2015).

⁵⁵ Council of Europe, *Algorithms and Human Rights*.

⁵⁶ Christian Sandvig and others, "Auditing algorithms: research methods for detecting discrimination on Internet platforms", paper presented at Data and Discrimination: Converting Critical Concerns into Productive Inquiry, a pre-conference at the 64th annual meeting of the International Communication Association, Seattle, 22 May 2014.

службы. Эта задача может решаться в рамках просветительских кампаний, использования всплывающих окон и других способов информирования о том, что система искусственного интеллекта определяет опыт пользователя во время работы в поисковой системе, посещения новостного сайта или платформы социальных сетей. В целях обеспечения информирования и получения согласия пользователя можно рассмотреть вопрос о целесообразности введения государством требования в отношении обязательного раскрытия информации. Частные лица также имеют право знать о том, осуществляет ли приложение искусственного интеллекта сбор личных данных и будут ли эти данные включены в набор данных, которые впоследствии будут определять работу интеллектуального приложения, и они также должны быть проинформированы об условиях использования, хранения и удаления этих данных.

60. Средства правовой защиты. Негативное воздействие систем искусственного интеллекта на права человека должно представлять собой исправимый ущерб, который должен быть устранен виновными в его причинении компаниями. Непременным условием создания эффективной системы правовой защиты является информирование частных лиц о том, что они были объектом алгоритмического решения (в том числе решения, предложенного системой искусственного интеллекта и утвержденное участником интерактивного процесса) и предоставление им информации о логике, лежащей в основе этого решения. Кроме того, компании должны обеспечить рассмотрение человеком ходатайств о предоставлении средств правовой защиты в целях надлежащей проверки систем и обеспечения подотчетности. Необходимо публиковать данные о числе случаев использования механизмов правовой защиты применительно к решениям, принятым с использованием технологий искусственного интеллекта.

V. Выводы и рекомендации

61. В настоящем докладе я изучил существующие и потенциальные последствия использования искусственного интеллекта для осуществления права на свободу мнений и их свободное выражение, утверждая что в настоящее время искусственный интеллект является важнейшим элементом информационной среды и создает выгоды и риски для лиц в рамках реализации своих прав. Я предложил концептуальную основу для размышлений относительно обязательств государств и обязанностей компаний, принятых с целью обеспечения соблюдения этих прав в условиях расширения технологических возможностей, и сформулировал конкретные меры, которые могут быть приняты правительствами и компаниями в целях обеспечения уважения прав человека в контексте расширения влияния, сферы охвата и масштабов применения технологий искусственного интеллекта.

Рекомендации для государств

62. В процессе осуществления закупок и развертывания систем искусственного интеллекта, государствам следует обеспечить, чтобы органы государственного сектора действовали в соответствии с правозащитными принципами. Эта работа, в частности включает, проведение консультаций с общественностью и подготовку оценок воздействия на права человека и разработку государственным учреждением оценок алгоритмического воздействия до осуществления закупок и развертывания системы искусственного интеллекта. Особое внимание следует уделить непропорциональному воздействию таких технологий на расовые и религиозные меньшинства, политическую оппозицию и активистов. Развертывание правительствами

систем искусственного интеллекта должно подлежать регулярной аудиторской проверке со стороны внешних независимых экспертов.

63. Государства должны обеспечить, чтобы права человека занимали центральное место в процессе разработки, развертывания и применения систем искусственного интеллекта компаниями частного сектора. Эта деятельность включает обновление и применение существующих правил в области искусственного интеллекта, в частности правил защиты данных, внедрение схем самостоятельного и совместного правового регулирования, призванных обеспечить подготовку компаниями оценок воздействия и проведение аудиторских проверок технологий искусственного интеллекта и создание эффективных внешних механизмов отчетности⁵⁷. При необходимости можно рассмотреть вопрос о целесообразности введения секторального регулирования конкретных приложений искусственного интеллекта в качестве необходимой и эффективной меры защиты прав человека. Если такие меры ведут к ущемлению и способствуют ограничению права на свободу выражения мнений, государства должны обеспечить, чтобы их действие было необходимым и соразмерным задаче достижения законной цели в соответствии со статьей 19 (3) Пакта. На основе широких консультаций с общественностью с участием представителей гражданского общества, правозащитных групп и маргинализованных или недопредставленных конечных пользователей необходимо разработать правовые нормы регулирования использования искусственного интеллекта.

64. Государствам следует создавать политические и законодательные условия, способствующие формированию среды для разнообразной и плюралистической информации. Решение этой задачи предусматривает принятие мер в целях обеспечения конкуренции в области искусственного интеллекта. Такие меры могут включать регулирование деятельности монополий в целях предотвращения концентрации технологии искусственного интеллекта и власти в руках нескольких господствующих на рынке компаний, разработку положений, призванных повысить уровень функциональной совместимости служб и технологий, а также принятие стратегии обеспечения нейтральности сети и оборудования⁵⁸.

Рекомендации для компаний

65. Все усилия по разработке руководящих принципов и кодексов об этических последствиях применения технологий искусственного интеллекта должны основываться на правозащитных принципах. В процессе разработки и развертывания любых систем искусственного интеллекта в частном и государственном секторах необходимо обеспечить гражданскому обществу возможность высказать свои замечания. Компаниям следует подтвердить в корпоративной политике и технических руководствах для разработчиков, инженеров, специалистов по обработке и очистке данных, программистов и других лиц, связанных с жизненным циклом искусственного интеллекта, что вся их деятельность осуществляется в соответствии с правозащитными обязательствами, и что этические принципы могут способствовать осуществлению правозащитных принципов в конкретных ситуациях, связанных с проектированием, внедрением и использованием систем

⁵⁷ Wagner, "Ethics as an escape from regulation".

⁵⁸ Autorité de régulation des communications électroniques et des postes, *Devices, the Weak Link in Achieving an Open Internet* (2018). URL: https://www.arcep.fr/uploads/tx_gspublication/rapport-terminaux-fev2018-ENG.pdf.

искусственного интеллекта. В частности, разработка положений об условиях предоставления услуг должна осуществляться на основе универсальных принципов в области прав человека.

66. Компании должны предоставить конкретную информацию о сферах и способах использования платформами, службами и прикладными программами технологий искусственного интеллекта и автоматизированных методов. Использование новаторских способов информирования пользователей о случаях, когда они участвуют в процессах принятия решений, принимаемых с помощью искусственного интеллекта, когда искусственный интеллект используется в целях представления или модерации контента или когда личные данные могут включаться в массивы данных, которые будут использоваться при разработке системы искусственного интеллекта, имеет решающее значение для предоставления в распоряжение пользователей информации, необходимой для понимания и преодоления негативных последствий воздействия систем искусственного интеллекта для осуществление прав человека. Компании должны публиковать данные об удалении контента, в том числе о количестве решений об удалении контента, которые становятся предметом споров и случаев успешно оспоренных решений, а также сведения о тенденциях в области отображения контента, тематических исследованиях и научных разработках по вопросам коммерческого и политического профилирования.

67. Компании должны предотвращать и нести ответственность за дискриминацию на уровне как входных, так и выходных данных систем искусственного интеллекта. Решение этой задачи предполагает использование группами специалистов по проектированию и развертыванию систем искусственного интеллекта разнообразных и недискриминационных представлений и предотвращение предвзятости и дискриминации при выборе массивов данных и разработке систем, в том числе путем устранения ошибок в выборке, очистки данных и наборов данных в целях удаления дискриминационных данных и принятия мер по компенсации таких данных. Важное значение имеет активный мониторинг дискриминационных решений, принимаемых системами искусственного интеллекта.

68. В ходе разработки и развертывания новых систем искусственного интеллекта, включая развертывание существующих систем на новых глобальных рынках, необходимо готовить оценки воздействия на права человека и проводить консультации с общественностью. До завершения разработки или внедрения нового продукта или службы необходимо провести публичные консультации и заручиться поддержкой общественности в целях проведения конструктивного диалога при активном участии представителей гражданского общества, правозащитников и представителей маргинализированных и недопредставленных конечных пользователей. Результаты оценок воздействия на права человека и консультаций с общественностью должны быть также обнародованы.

69. Компании должны обеспечить аудирование всех алгоритмов кода искусственного интеллекта и использовать новаторские формы проведения внешней независимой аудиторской проверки системы искусственного интеллекта в дополнение к ревизиям, проводимым согласно требованиям действующего регламента. Результаты аудиторских проверок искусственного интеллекта также должны быть обнародованы.

70. Индивидуальные пользователи должны иметь доступ к средствам правовой защиты от негативного воздействия систем искусственного интеллекта на права человека. Компаниям следует создать системы

рассмотрения человеком жалоб пользователей, предоставления средств правовой защиты и оперативного рассмотрения апелляций на решения, принятые системами искусственного интеллекта. Необходимо на регулярной основе публиковать данные о числе жалоб на решения систем искусственного интеллекта и ходатайств о предоставлении средств правовой защиты и формах и эффективности имеющихся средств правовой защиты.
