

**Совет по правам человека****Пятьдесят шестая сессия**

18 июня — 14 июля 2024 года

Пункт 9 повестки дня

**Расизм, расовая дискриминация, ксенофобия
и связанные с ними формы нетерпимости:****последующие меры и осуществление****Дурбанской декларации и Программы действий****Современные формы расизма, расовой дискриминации,
ксенофобии и связанной с ними нетерпимости****Доклад Специального докладчика по вопросу о современных
формах расизма, расовой дискриминации, ксенофобии и связанной
с ними нетерпимости Ашвини К.П.****Резюме*

В настоящем докладе Специальный докладчик по вопросу о современных формах расизма, расовой дискриминации, ксенофобии и связанной с ними нетерпимости Ашвини К.П. подводит итоги своей деятельности за прошедший год и рассматривает вопрос о том, как господствующее представление об объективности и нейтральности технологий позволяет искусственному интеллекту увековечивать расовую дискриминацию. Она рассматривает четыре сквозных канала, по которым искусственный интеллект может способствовать проявлениям расовой дискриминации: проблемы с данными, проблемы разработки алгоритмов, намеренное дискриминационное использование искусственного интеллекта и вопросы подотчетности. Затем она приводит примеры применения искусственного интеллекта в различных сферах жизни общества и его последствий, связанных с расовой дискриминацией. Она анализирует нарождающееся стремление управлять искусственным интеллектом и регулировать его применение, а затем делает обзор соответствующих международных стандартов в области прав человека. В заключение она представляет рекомендации, как государствам следует подходить к управлению использованием технологий искусственного интеллекта и его регулированию для предотвращения и искоренения расовой дискриминации.

* Была достигнута договоренность о публикации настоящего документа после стандартной даты публикации по независящим от составителя обстоятельствам.



I. Введение

1. Настоящий доклад представляется в соответствии с резолюцией 52/36 Совета по правам человека, в которой Совет просил Специального докладчика по вопросу о современных формах расизма, расовой дискриминации, ксенофобии и связанной с ними нетерпимости представлять ему ежегодный доклад. В докладе Специальный докладчик описывает свою деятельность в рамках мандата и рассматривает тему искусственного интеллекта (ИИ) и расовой дискриминации.
2. В связи с подготовкой доклада Специальный докладчик обратилась к государствам — членам Организации Объединенных Наций и другим заинтересованным сторонам, включая организации гражданского общества, международные организации и национальные правозащитные учреждения, с просьбой представить материалы. Специальный докладчик выражает искреннюю благодарность всем государствам-членам и другим заинтересованным сторонам, представившим информацию. Она использовала полученные материалы при подготовке настоящего доклада и остается открыта для продолжения диалога со всеми соответствующими заинтересованными сторонами по этой важной теме¹.

II. Краткое описание деятельности

3. В октябре 2023 года Специальный докладчик представила Генеральной Ассамблее на ее семьдесят восьмой сессии свои доклады о борьбе с героизацией нацизма, неонацизмом и другими видами практики, которые способствуют эскалации современных форм расизма, расовой дискриминации, ксенофобии и связанной с ними нетерпимости, и о распространении расистских ненавистнических высказываний в Интернете². В период с 31 октября по 14 ноября 2023 года Специальный докладчик совершила поездку в Соединенные Штаты Америки³.
4. В августе 2023 года Специальный докладчик приняла участие в девятой сессии Группы независимых видных экспертов по вопросу об осуществлении Дурбанской декларации и Программы действий. В январе 2024 года Специальный докладчик приняла участие в региональном совещании для Азиатско-тихоокеанского региона, посвященном Международному десятилетию лиц африканского происхождения. В феврале 2024 года она приняла участие в Международной конференции по продовольственной справедливости под углом зрения прав человека на тему «Вызовы реальности и ставки на будущее», которая проходила в Катаре. В апреле 2024 года она приняла участие в третьей сессии Постоянного форума по лицам африканского происхождения, где выступила с докладом о преодолении системного расизма и исторического вреда в сфере образования.

III. Искусственный интеллект и расовая дискриминация

5. В настоящем докладе Специальный докладчик решила сосредоточиться на искусственном интеллекте и расовой дискриминации. Эта тема согласуется с тем стратегическим акцентом, который она делает на взаимосвязи между цифровыми технологиями и расовой дискриминацией, о чем говорится в ее докладе пятьдесят третьей сессии Совету по правам человека, где изложены ее стратегическое видение и первоначальные приоритеты⁴. Она опирается на работу предыдущего мандатария,

¹ Специальный докладчик получала исследовательскую и аналитическую поддержку от Международной клиники прав человека Гарвардской школы права и Международной клиники прав человека и разрешения конфликтов и Стэнфордского центра расовой справедливости Стэнфордской школы права. Она искренне благодарит всех, кто принимал участие в подготовке доклада, за неоценимую поддержку.

² A/78/302 и A/78/538.

³ См. A/HRC/56/68/Add.1.

⁴ A/HRC/53/60, пп. 50–53.

занимавшегося вопросами новых цифровых технологий и расовой дискриминации⁵, и реагирует на тот интерес, который Совет по правам человека и вся система Организации Объединенных Наций проявляют к управлению использованием искусственного интеллекта⁶.

6. В связи с последними разработками в области генеративного искусственного интеллекта и все более широким применением искусственного интеллекта продолжают возникать серьезные вопросы, касающиеся прав человека, в том числе озабоченность по поводу расовой дискриминации. Генеративный искусственный интеллект меняет мир и в будущем он может стать двигателем все более масштабных изменений в обществе. Стремительное расширение использования искусственного интеллекта в различных областях вызывает глубокую озабоченность Специального докладчика. Это не означает, что искусственный интеллект не сулит потенциальных выгод. На самом деле, он открывает широкие возможности для инноваций и инклюзивности. Однако такие технологии растут и развиваются практически безудержными темпами. Специальный докладчик обеспокоен тем, что политические и правовые меры, направленные на управление использованием искусственного интеллекта и его регулирование, не успевают за развитием этой технологии и что первые попытки наладить такое управление и регулирование в недостаточной мере учитывают огромную потенциальную роль этой технологии, как нынешнюю, так и будущую, в сохранении и углублении системной расовой дискриминации, а также в усилении неравенства внутри регионов, стран и общин и между ними.

7. Как отметил предыдущий мандатарий, существует устойчивое и вредное представление о том, что технология нейтральна и объективна:

Общественное восприятие технологии обычно сводится к тому, что технология по своей природе нейтральна и объективна, причем некоторые отмечают, что такая презумпция технологической объективности и нейтральности остается явно выраженной даже среди производителей технологии. Но технология никогда не бывает нейтральной — она отражает ценности и интересы тех, кто влияет на ее разработку и использование, и в ее основе лежат те же самые структуры неравенства, которые действуют в обществе⁷.

8. В настоящем докладе Специальный докладчик рассматривает вопрос о том, каким образом господствующие представления об объективности и нейтральности технологий позволяют искусственному интеллекту увековечивать расовую дискриминацию.

A. Сквозные каналы, по которым искусственный интеллект может способствовать проявлениям расовой дискриминации

9. Искусственный интеллект не является монолитом. Действительно, существует несколько его типов. Предиктивный искусственный интеллект считается «традиционной» формой технологии, и его модели используют исторические данные, закономерности и тенденции, чтобы делать обоснованные прогнозы относительно будущих событий или результатов.

10. Искусственный интеллект, используемый для идентификации печатных символов, человеческих лиц, объектов и другой информации, является еще одной формой «традиционного» искусственного интеллекта и включает в себя различные технологии распознавания и различения объектов, лиц и закономерностей в данных, которые ему предоставляются.

⁵ См. A/75/590, A/HRC/44/57 и A/HRC/48/76.

⁶ См., например, High-level Advisory Body on Artificial Intelligence, “Interim report: governing AI for humanity” (December 2023); резолюция 53/29 Совета по правам человека; и резолюции 78/213 и 78/265 Генеральной Ассамблеи.

⁷ A/HRC/44/57, п. 12.

11. Генеративные системы искусственного интеллекта — это новые формы искусственного интеллекта. Они универсальны и могут использоваться для самых разных целей. Они относятся к классу систем искусственного интеллекта, предназначенных для получения разнообразных результатов на основе обширных наборов обучающих данных, нейронных сетей, архитектуры глубокого обучения и подсказок для пользователей. Генеративные модели искусственного интеллекта могут создавать на выходе широкий спектр данных, включая изображения, текст, аудио, видео и синтетические данные. В отличие от моделей искусственного интеллекта, ориентированных на выявление закономерностей в существующих данных, генеративный искусственный интеллект обучается создавать новые единицы данных, которые имитируют закономерности и характеристики данных, используемых для обучения моделей машинного обучения. Появление генеративного искусственного интеллекта приведет к появлению множества новых приложений, а также к возникновению множества новых вопросов, связанных с правами человека⁸.

12. Эти различные виды искусственного интеллекта имеют множество применений. Более конкретные примеры использования искусственного интеллекта и связанные с этим последствия с точки зрения расовой дискриминации Специальный докладчик рассматривает ниже. Однако она хотела бы подчеркнуть важность изучения общих закономерностей, позволяющих оценить, как искусственный интеллект может увековечить расовую дискриминацию, особенно в рамках правовых и политических дебатов, связанных с управлением использованием искусственного интеллекта и его регулированием. В ходе таких дебатов последствия использования искусственного интеллекта необходимо рассматривать под призмой системного расизма, определяемого как «действие сложной, взаимосвязанной системы законов, политики, практики и взглядов в государственных учреждениях, частном секторе и общественных структурах, которые в совокупности приводят к прямой или косвенной, преднамеренной или непреднамеренной, де-юре или де-факто дискриминации, различию, исключению, ограничению или предпочтению по признаку расы, цвета кожи, родового, национального или этнического происхождения»⁹. Как следует из этого определения, системный расизм — это сложное, зачастую ползучее и распространенное в обществе явление. Проявления системного расизма в одной области являются взаимосвязанными, взаимозависимыми и взаимно усиливающими по отношению к его проявлениям в других областях. Рассмотрение сквозных каналов, по которым искусственный интеллект способствует расовой дискриминации, может помочь определить, как он взаимодействует с проявлениями системного расизма и усиливает их, а также в целом усиливает системное угнетение в обществе по расовому и этническому признакам¹⁰.

1. Проблемы с данными

13. Развитие систем искусственного интеллекта и алгоритмов машинного обучения привело к масштабной оцифровке данных. Алгоритмы используют эти данные для принятия решений и осуществления действий в различных отраслях. Однако наборы данных, на которых обучаются алгоритмы, зачастую неполны или недостаточно точно представляют определенные группы людей. Если в обучающих наборах данных определенные группы пере- или недопредставлены, в том числе по расовому и этническому признакам, это может привести к искажению алгоритмов. Аналогичным образом, если обучающие наборы уже содержат искаженные данные, они могут дать искаженные результаты.

14. Если обучающих данных недостаточно, построенные на основе алгоритмов прогнозы могут систематически дискриминировать группы, которые не представлены или недостаточно представлены в данных. Искажение алгоритмов возможно не только при слишком малом количестве данных; алгоритмы, основанные на

⁸ Материалы, представленные Австралийской комиссией по правам человека. Все представленные материалы будут размещены на веб-сайте Управления Верховного комиссара Организации Объединенных Наций по правам человека (УВКПЧ).

⁹ A/HRC/47/53, п. 9.

¹⁰ A/HRC/44/57, п. 43.

нерепрезентативных данных, также могут давать искаженные результаты. Например, исследование, посвященное базам видеоданных правоохранительных органов США, показало, что люди африканского происхождения чаще других ошибочно идентифицируются как нарушители в сетях распознавания лиц, используемых сотрудниками правоохранительных органов. Это объясняется ошибками в идентификации лиц этой группы и перепредставленностью лиц африканского происхождения в полицейских базах фотоснимков, что объясняется историческим наследием системного расизма¹¹.

15. Исторические предубеждения могут влиять на сами данные. Основным элементом машинного обучения является составление прогнозов на будущее на основе данных из прошлого. Однако если прошлые данные необъективны по отношению к определенным группам, в том числе по расовому и этническому признакам, компьютерные модели могут воспроизводить и усиливать эту необъективность. Использование необъективных или искаженных данных при принятии решений в реальной жизни может нанести еще больший вред маргинализированным расовым и этническим группам, поскольку использование таких данных в контексте искусственного интеллекта ведет к созданию еще большего объема данных, которые затем используются для принятия будущих решений. Такие самовоспроизводящие системы могут повторять и углублять существующие искажения.

16. Последняя проблема, связанная с данными, — это проблема конфиденциальности. Данные, используемые в системах искусственного интеллекта, часто содержат личную информацию людей, к которым эти данные относятся. Сбор и обработка данных без согласия нарушают право на неприкосновенность частной жизни. Бывают также случаи, когда данные, собранные в одних целях, например в системе здравоохранения, в том числе с помощью медицинских приложений, без согласия передаются для использования в других, например в правоохранительных целях. Утечки данных и несанкционированный доступ к личной информации в результате взлома создают дополнительные проблемы с конфиденциальностью. Для представителей расово маргинализированных групп проблемы прав человека, связанные с правом на частную жизнь, могут стоять острее. Нарушение конфиденциальности может подвергнуть эти группы риску остракизма, дискриминации или физической опасности¹².

2. Проблемы алгоритмического дизайна

17. Вторая распространенная причина искажений в работе инструментов искусственного интеллекта связана с тем, как разрабатываются алгоритмы. Если искажения связаны с выбором дизайна, алгоритм может способствовать получению необъективных результатов, даже если данные, вводимые в алгоритм, идеально репрезентативны. Причины искажений могут крыться в решениях о параметрах и функционировании алгоритма. Разработчики алгоритмов принимают решения о том, какие переменные будут использоваться в алгоритме, как определить категории или пороговые значения для сортировки информации и какие данные будут использоваться для построения алгоритма. Выбор разработчиков касается также измерения конкретных характеристик и критериев успеха использования алгоритма. Иногда биография или взгляды разработчиков алгоритмов могут привести к тому, что в алгоритмы будут неосознанно заложены их предубеждения, в том числе расовые¹³. Отмечается, что недостаточное разнообразие в секторе цифровых технологий усугубляется отсутствием инклюзивных процессов консультаций при разработке

¹¹ Nicol Turner Lee, Paul Resnick and Genie Barton, “Algorithmic bias detection and mitigation: best practices and policies to reduce consumer harms”, Brookings Institution, 22 May 2019.

¹² Samantha Lai and Brooke Tanner, “Examining the intersection of data privacy and civil rights”, Brookings Institution, 18 July 2022. См. также материалы, представленные Privacy International.

¹³ Ninareh Mehrabi and others, “A survey on bias and fairness in machine learning”, *ACM Computing Surveys*, vol. 54, No. 6 (2022); материалы, представленные The London Story; и A/HRC/44/57, п. 17.

систем искусственного интеллекта, что способствует возникновению проблем с конструированием алгоритмов¹⁴.

18. Выбор алгоритмического дизайна может иметь серьезные дискриминационные последствия в реальной жизни. Например, при разработке алгоритма оценки кредитного риска способ определения и измерения «риска» может привести к дискриминационным результатам. Если разработчик алгоритма решит использовать кредитные рейтинги в качестве косвенного показателя риска, это может привести к дискриминационным последствиям для групп людей, которые, как правило, имеют более низкие кредитные рейтинги. Исследования показывают, что между кредитными рейтингами, расой и другими демографическими показателями существует сильная корреляция и что использование кредитных рейтингов ставит в невыгодное положение определенные группы населения¹⁵. Эта корреляция во многих случаях может рассматриваться как побочный продукт существующего системного расизма и отчуждения. Люди могут оказаться в невыгодном положении из-за того, что разработчик алгоритма решил использовать кредитный рейтинг для оценки кредитного риска, хотя такой рейтинг формально не является дискриминационным критерием.

3. Использование в дискриминационных целях

19. В некоторых случаях искусственный интеллект может использоваться в откровенно расистских целях, избирательно применяясь против целевых групп, что приводит к дискриминационным результатам. Например, поступают сообщения о том, что правоохранительные органы намеренно используют искусственный интеллект для слежки и чрезмерного полицейского контроля за конкретными сообществами по расовому признаку¹⁶. Кроме того, преднамеренная дискриминация может иметь место, когда правительства и другие субъекты используют возможности технологий для мониторинга, составления профиля и целенаправленной обработки групп или отдельных лиц по признаку их расовой или этнической принадлежности¹⁷.

20. Распространение дезинформации — еще один способ использования искусственного интеллекта в откровенно расистских целях. Политические деятели могут использовать искусственный интеллект для создания текстов, изображений и видеороликов, чтобы манипулировать общественным мнением и политическими процессами в свою пользу и подрывать доверие к институтам, в том числе по расовому признаку. Сообщается также, что правительства используют искусственный интеллект, чтобы сеять раздор и насаждать цензуру в Интернете¹⁸.

4. Проблемы подотчетности

21. Тот факт, что некоторые инструменты искусственного интеллекта принимают решения независимо от человека, означает, что процесс принятия решений скрыт, словно в непроницаемом «черном ящике». Кроме того, алгоритм может принимать решения самостоятельно, поскольку, получая данные, алгоритмы искусственного интеллекта постоянно обновляют себя. Со временем инструмент искусственного интеллекта может при принятии решений начать опираться на факторы, на которые он изначально не был запрограммирован. В основу этих факторов лягут закономерности, которые искусственный интеллект сам выявит в данных. По мере того как алгоритм будет включать эти выявленные закономерности в свой код и принимать решения на их основе, люди, полагающиеся на алгоритм, уже не смогут «заглянуть под капот» и точно определить критерии, которые алгоритм использовал для получения

¹⁴ Материалы, представленные NetMission.Asia.

¹⁵ A.R. Lange and Natasha Duarte, “Understanding bias in algorithmic design”, Medium, 6 September 2017.

¹⁶ См. Amnesty International, *Decode Surveillance NYC: Methodology* (London, 2022); и материалы, представленные NetMission.Asia.

¹⁷ Материалы, представленные NetMission.Asia.

¹⁸ Tate Ryan-Mosley, “How generative AI is boosting the spread of disinformation and propaganda”, *MIT Technology Review*, 4 October 2023.

определенных результатов. Таким образом, проблема «черного ящика» делает логический процесс искусственного интеллекта непредсказуемым и непрозрачным¹⁹. Кроме того, многие алгоритмы, разработанные коммерческими организациями, не могут быть проверены из-за законов о защите контрактов и интеллектуальной собственности, что усугубляет проблему подотчетности²⁰.

22. Проблема «черного ящика» имеет особенно серьезные последствия в контексте системного расизма. Как уже говорилось выше, системный расизм — это не бросающееся в глаза, но глубоко разрушительное зло в масштабах всего общества. Силы, движущие системным расизмом, не всегда признаются, и это усугубляется сохраняющимися пробелами в собираемых данных, дезаггированных по расовому и этническому признаку²¹. Искусственный интеллект без эффективных механизмов подотчетности способен стать дополнительной движущей силой уже существующего разрушительного явления ползучего системного расизма.

23. Проблемы подотчетности в сфере искусственного интеллекта серьезно влияют на возможность тех, кто подвергается расовой дискриминации, получать эффективную правовую защиту. Сегодня, когда представители маргинализированных расовых и этнических групп сталкиваются с различными результатами из-за принимаемых людьми решений, суды и другие механизмы обеспечения подотчетности могут проверить, были ли эти действия преднамеренными и оправданными²². Когда решения принимают люди, часто есть доказательства, которые можно использовать для таких проверок. Во многих случаях автономные процессы принятия решений не оставляют доказательств, как это делают люди, принимающие решения²³. Проблемы «черного ящика» усиливают и без того серьезные препятствия для доступа к правосудию для тех, кто подвергается расовой дискриминации.

В. Использование искусственного интеллекта и его дискриминационные последствия

24. В настоящем разделе Специальный докладчик приводит примеры применения искусственного интеллекта в различных сферах жизни общества и его последствий, связанных с расовой дискриминацией. Эти примеры являются иллюстративными и неисчерпывающими и приводятся в качестве наглядного доказательства того, что искусственный интеллект уже способствует расовой дискриминации. Специальный докладчик рассматривает эти примеры как взаимосвязанные и взаимоусиливающие проявления расовой дискриминации, которые способствуют всестороннему усилению системного угнетения по расовому и этническому признакам в масштабах общества.

25. Специальный докладчик выбрал три области для демонстрации дискриминационного воздействия искусственного интеллекта: правоохранительная деятельность, безопасность и система уголовного правосудия; образование; и здравоохранение. Что касается использования искусственного интеллекта в других контекстах, Специальный докладчик рекомендует ознакомиться с докладами предыдущего мандатария о появлении цифровых границ и фиксации случаев расовой и ксенофобной дискриминации в контексте цифровых процедур пограничного и иммиграционного контроля, а также об использовании цифровых технологий в пограничных и иммиграционных службах²⁴. Специальный докладчик также отсылает читателей к своему докладу семьдесят восьмой сессии Генеральной Ассамблеи о ненавистнических высказываниях на почве расизма в Интернете, в котором

¹⁹ Yavar Bathaee, “The artificial intelligence black box and the failure of intent and causation”, *Harvard Journal of Law and Technology*, vol. 31, No. 2 (2018); A/HRC/44/57, п. 34; и Renata M. O’Donnell, “Challenging racist predictive policing algorithms under the Equal Protection Clause”, *New York University Law Review*, vol. 94, No. 3 (June 2019).

²⁰ A/HRC/44/57, п. 44.

²¹ A/HRC/47/53, п. 16.

²² Bathaee, “The artificial intelligence black box”.

²³ Там же.

²⁴ A/75/590 и A/HRC/48/76.

рассматривается использование искусственного интеллекта для цензурирования контента в социальных сетях²⁵, и к докладу Специального докладчика по вопросу о крайней нищете и правах человека семьдесят четвертой сессии Генеральной Ассамблеи, в котором содержится анализ использования искусственного интеллекта в системах социальной защиты²⁶.

1. Правоохранительная деятельность, безопасность и система уголовного правосудия

а) Автоматизированная идентификация

26. Правоохранительные органы используют автоматизированные инструменты идентификации, чтобы связать то, что они наблюдают в конкретной обстановке, с потенциальным «совпадением» в базе данных. Одним из наиболее распространенных видов автоматизированных средств идентификации является технология распознавания лиц. Средства распознавания лиц получают видеозаписи или фотографии человека и вводят их в алгоритмы. Затем алгоритмы сравнивают изображения с базой данных полицейских фотографий, фотографий водительских прав или других изображений с целью идентификации человека²⁷. Разработчики таких инструментов обучают модели, которые лежат в их основе, показывая им изображения лиц в процессе машинного обучения. Цель — обучить модели определять отличительные особенности человеческих лиц²⁸. Однако наборы изображений, используемые для обучения этих моделей, не всегда являются демографически репрезентативными²⁹. В одном из исследований, посвященном популярной базе данных, содержащей изображения, ученые обнаружили перепредставленность мужчин в возрасте от 18 до 40 лет и недопредставленность людей с темной кожей³⁰. Согласно другому исследованию, посвященному коммерческим системам распознавания лиц, алгоритмы гендерной классификации обучаются на наборах данных с подавляющим большинством белых мужских лиц³¹. Отсутствие расового, гендерного и культурного разнообразия в наборах данных для обучения искусственного интеллекта приводит к одной из классических проблем с данными, описанных выше. Для групп, недостаточно представленных в обучающих данных, в том числе для тех, которые подвергаются межсекторальной дискриминации, алгоритм с большей вероятностью обнаружит неверное совпадение.

27. Сообщалось, что неправильная идентификация лиц с помощью этих технологий привела к увеличению числа арестов людей африканского происхождения³². Специальный докладчик по вопросу о поощрении и защите права на свободу мнений и их свободное выражение и Верховный комиссар Организации Объединенных Наций по правам человека отметили, что средства распознавания лиц часто способствуют незаконной дискриминации и расовому профилированию³³. Несмотря на такие опасения по поводу прав человека, системы распознавания лиц используются правоохранительными органами в ряде стран. Например, сообщается, что значительные средства в такие системы инвестировало правительство Индии. Система распознавания лиц, используемая полицией Дели, по сообщениям, была точной только

²⁵ A/78/538.

²⁶ A/74/493.

²⁷ Marissa Gerchick and Matt Cagle, “When it comes to facial recognition, there is no such thing as a magic number”, American Civil Liberties Union, 7 February 2024.

²⁸ Julia Dressel and Andrew Warren, “Breaking down data analytics and AI in criminal justice”, Recidiviz, 8 March 2022.

²⁹ Материалы, представленные AI for the People.

³⁰ Khari Johnson, “ImageNet creators find blurring faces for privacy has a ‘minimal impact on accuracy’”, VentureBeat, 16 March 2021.

³¹ Joy Buolamwini and Timnit Gebru, “Gender shades: intersectional accuracy disparities in commercial gender classification”, *Proceedings of Machine Learning Research*, vol. 81 (2018). См. также Gerchick and Cagle, “When it comes to facial recognition, there is no such thing as a magic number”; материалы, представленные AI for the People; и материалы, представленные Internet Lab.

³² Gerchick and Cagle, “When it comes to facial recognition, there is no such thing as a magic number”.

³³ См. A/HRC/41/35 и A/HRC/48/31.

в 2 % случаев и подвергала представителей меньшинств непропорционально высокому риску неправильной идентификации и ложного ареста³⁴. По сообщениям, бразильские правоохранительные органы ложно обвиняют и арестовывают людей, опираясь на неточные инструменты распознавания лиц. Согласно исследованию, проведенному в 2019 году, 90 % людей, арестованных в бразильских городах на основе технологии распознавания лиц, имеют африканское происхождение³⁵.

28. Системы засечки выстрелов — еще один распространенный тип автоматизированных средств идентификации, используемых сотрудниками правоохранительных органов в ряде стран. Одна из систем, получившая название ShotSpotter, предполагает установку в жилых районах датчиков с микрофоном, снабженных системой GPS, памятью и возможностями обработки данных, а также подключенных к сотовой связи³⁶. Когда датчики улавливают шум, который может быть выстрелом, алгоритм триангулирует местоположение источника шума. Алгоритм отфильтровывает другие возможные звуки, прежде чем отправить аудиозапись человеку для прослушивания³⁷. Имеющаяся информация свидетельствует о том, что системы засечки выстрелов непропорционально часто используются в районах, населенных расово маргинализированными группами, и погрешность этих устройств может быть очень высокой³⁸. Установка систем засечки выстрелов в районах, где проживают маргинализированные расовые и этнические группы, и неточности таких систем усугубляют системные предубеждения среди сотрудников правоохранительных органов.

29. Существует множество примеров того, как использование технологии автоматической идентификации имело судьбоносные последствия. В 2019 году в США чернокожий мужчина был ошибочно арестован в Нью-Джерси и просидел в тюрьме 10 дней из-за ошибки в распознавании лиц. Несмотря на наличие доказательств его невиновности, власти почти год не прекращали его дело, и за выдвинутые против него обвинения ему грозило до 25 лет тюремного заключения. Этот инцидент оказал сильное влияние на жизнь мужчины³⁹. В феврале 2024 года сотрудники правоохранительных органов Чикаго, отреагировав на ложное предупреждение ShotSpotter, открыли огонь по ребенку, запускавшему фейерверк⁴⁰. Еще одним примером использования технологий искусственного интеллекта такого типа является внедрение Армией обороны Израиля системы Wolf Pack — обширной базы данных, содержащей визуальную и всю другую доступную информацию о палестинцах с Западного берега реки Иордан, в которую также интегрированы различные программы наблюдения, такие как Blue Wolf и Red Wolf⁴¹. По сообщениям, в Старом городе Хеврона Армия обороны Израиля установила камеры с искусственным интеллектом, способные распознавать человеческие лица, которые подключены к программе Blue Wolf — мобильному приложению, позволяющему военнослужащим обнаруживать и классифицировать палестинцев на Западном берегу с помощью обширной биометрической базы данных, собранной в большинстве случаев без согласия палестинцев, в результате чего за ними ведется постоянное слежение. Активное применение системы Wolf Pack Армией обороны Израиля

³⁴ Amnesty International, “Ban the scan: Hyderabad”, available at <https://banthescan.amnesty.org/hyderabad/>.

³⁵ Материалы, представленные группой экспертов из Бразилии.

³⁶ Alisha Ebrahimji, “Critics of ShotSpotter gunfire detection system say it’s ineffective, biased and costly”, CNN, 24 February 2024.

³⁷ Jay Stanley, “Four problems with the ShotSpotter gunshot detection system”, American Civil Liberties Union, 24 August 2021.

³⁸ Там же; и MacArthur Justice Center, “ShotSpotter is deployed overwhelmingly in Black and Latinx neighborhoods in Chicago”, available at <https://endpolicesurveillance.com/burden-on-communities-of-color/>.

³⁹ Gerchick and Cagle, “When it comes to facial recognition, there is no such thing as a magic number”; и Khari Johnson, “How wrongful arrests based on AI derailed 3 men’s lives”, *Wired*, 7 March 2022.

⁴⁰ Adam Schwartz, “Responding to ShotSpotter, police shoot at child lighting fireworks”, Electronic Frontier Foundation, 22 March 2024.

⁴¹ Amnesty International, *Automated Apartheid: How Facial Recognition Fragments, Segregates and Controls Palestinians in the OPT* (London, 2023), pp. 41–45.

усугубляет апартеид, жертвами которого становятся палестинцы⁴². Эти примеры свидетельствуют о серьезном влиянии на ситуацию с правами человека в результате использования систем искусственного интеллекта для принятия важных решений в условиях повышенного риска.

в) Алгоритмы предиктивного полицейского контроля

30. Еще одна форма технологии искусственного интеллекта, которая широко используется правоохранительными органами, — это предиктивный полицейский контроль. На основе данных о местонахождении и личных данных инструменты предиктивного полицейского контроля делают выводы о том, кто совершит преступления в будущем и где они могут произойти.

31. Предиктивный полицейский контроль может усугубить исторически сложившуюся чрезмерную полицейскую опеку над общинами по расовому и этническому признакам⁴³. Поскольку сотрудники правоохранительных органов исторически уделяют особое внимание тем районам, где проживают эти общины, их представители слишком часто попадают в полицейские сводки. Это, в свою очередь, влияет на прогнозируемое алгоритмом место совершения будущего преступления, что приводит к увеличению числа полицейских в этих районах⁴⁴. Предиктивный полицейский контроль также может иметь отношение к проблеме «черного ящика», поскольку алгоритмам не хватает прозрачности, в том числе в отношении того, какие данные анализируются и как используются прогнозы⁴⁵.

32. Алгоритмы предиктивного полицейского контроля на основе местоположения используют связи между местами, событиями и историческими данными о преступности, чтобы предсказать, когда и где могут произойти преступления в будущем⁴⁶. Затем сотрудники полиции соответствующим образом планируют патрулирование. Когда в районах с чрезмерно высоким уровнем полицейской активности сотрудники полиции фиксируют новые правонарушения, возникает замкнутый круг, когда на основе полученной информации алгоритм генерирует все более предвзятые прогнозы, в которых эти районы называются особенно проблемными. Короче говоря, прошлые искажения приводят к искажениям в будущем. В Соединенном Королевстве Великобритании и Северной Ирландии проведенное по заказу правительства исследование алгоритмической предвзятости в работе полиции показало, что определение географических мест как «горячих точек» преступности заставляет полицейских ожидать еще большего количества преступлений в этих районах. В результате сотрудники полиции чаще останавливают или арестовывают людей в «горячих точках» скорее по причине предвзятого отношения, чем исходя из реальных потребностей общественной безопасности⁴⁷. В Уругвае исследователи обнаружили, что данные, используемые в алгоритмах предиктивного полицейского контроля на основе местоположения, могут быть необъективными. Переменная местоположения может выступать в качестве

⁴² Sophia Goodfriend, “Algorithmic State violence: automated surveillance and Palestinian dispossession in Hebron’s Old City”, *International Journal of Middle East Studies*, vol. 55, No. 3 (2023).

⁴³ Tim Lau, “Predictive policing explained”, Brennan Center for Justice, 1 April 2020; и Jon Fasman, “The black box of justice: how secret algorithms have changed policing”, *Fast Company*, 9 February 2021.

⁴⁴ Kristian Lum and William Isaac, “To predict and serve?”, *Significance*, vol. 13, No. 5 (2016); и материалы, представленные Австралийской комиссией по правам человека.

⁴⁵ Lau, “Predictive policing explained”.

⁴⁶ Will Douglas Heaven, “Predictive policing algorithms are racist. They need to be dismantled”, *MIT Technology Review*, 17 July 2020.

⁴⁷ Там же. См. также Government of the United Kingdom of Great Britain and Northern Ireland, “Report commissioned by CDEI calls for measures to address bias in police use of data analytics”, 16 September 2019.

косвенного показателя социально-экономического или этнического происхождения, провоцируя дискриминацию⁴⁸.

33. Инструменты предиктивного полицейского контроля на основе личных данных позволяют предсказать, кто может совершить преступление в будущем, исходя из биографических данных о человеке. Биографические данные могут включать в себя возраст, пол, семейное положение, историю злоупотребления психоактивными веществами и судимости. Как и в случае с инструментами, основывающимися на местоположении, сведения о прошлых арестах, которые часто омрачаются системным расизмом в системе уголовного правосудия, могут исказить будущие прогнозы этих алгоритмов. Такие переменные, как социально-экономическое положение, уровень образования и место проживания, могут выступать косвенными признаками расы и увековечивать исторические предубеждения⁴⁹. В Австралии полиция Нового Южного Уэльса использовала основанный на алгоритмах план адресной работы с подозреваемыми для выявления лиц, склонных к совершению уголовных преступлений. По имеющимся сведениям, до отмены этого плана количество контактов полиции с представителями общин аборигенов и жителей островов Торрессова пролива было непропорционально высоким⁵⁰.

с) Алгоритмы оценки рецидивизма

34. Инструменты оценки рецидивизма используются для принятия решений на разных этапах системы уголовного правосудия, в том числе решений о залоге, поручительстве, вынесении приговора и условно-досрочном освобождении⁵¹. Инструменты оценки рецидивизма используют исторические данные для оценки вероятности совершения обвиняемыми определенных действий, в частности вероятности совершения ими нового преступления в будущем. Эти инструменты позволяют составить рейтинги риска, используя информацию из таких источников, как уголовные дела и опросы обвиняемых⁵².

35. Инструменты прогнозирования рецидивизма демонстрируют многочисленные проблемы искусственного интеллекта, которые способствуют расовой дискриминации. Во-первых, у инструментов есть проблемы с данными. Данные системы уголовного правосудия, использованные для обучения алгоритмов, отражают системное неравенство, основанное на истории расистского поведения полиции⁵³. Кроме того, выбор дизайна, например способа измерения или оценки переменных, может способствовать алгоритмической дискриминации⁵⁴. Помимо этого, то, как разработчик алгоритма определяет «успех», может повлиять на то, что алгоритм стремится максимизировать. Если алгоритм настроен на оптимизацию для минимального количества новых преступлений, он может соотносить более длительные сроки заключения с более низким уровнем рецидива, поскольку люди не могут совершить повторное преступление, находясь в заключении. Поэтому он может использовать эти закономерности, чтобы рекомендовать более длительные сроки заключения.

36. Исследователи предполагают, что прогнозные показатели рецидивизма не являются точными и что их ошибки оказывают непропорционально большое влияние на расовые маргинальные группы. Например, исследование, проведенное в США,

⁴⁸ Juan Ortiz Freuler and Carlos Iglesias, “Algorithms and artificial intelligence in Latin America: a study of implementation by governments in Argentina and Uruguay”, World Wide Web Foundation, September 2018; и Фонд Eticas, Eticas Foundation, “Uruguay’s Ministry of the Interior invests in predictive policing”, 13 September 2021.

⁴⁹ Heaven, “Predictive policing algorithms are racist”.

⁵⁰ Материалы, представленные Австралийской комиссией по правам человека.

⁵¹ Julia Angwin and others, “Machine bias”, ProPublica, 23 May 2016.

⁵² Там же.

⁵³ См. Heaven, “Predictive policing algorithms are racist”; и Michael Mayowa Farayola and others, “Fairness of AI in predicting the risk of recidivism: review and phase mapping of AI fairness techniques”, in *Proceedings of the 18th International Conference on Availability, Reliability and Security* (Association for Computing Machinery, 2023).

⁵⁴ Mehrabi and others, “A survey on bias and fairness in machine learning”.

показало, что рейтинги риска очень ненадежны при прогнозировании насильственных преступлений. По имеющимся данным, людей африканского происхождения ошибочно считают будущими преступниками почти в два раза чаще, чем белых.

d) Автономные системы вооружений

37. Автономные системы вооружений включают в себя любые оружейные системы с автономным управлением важнейшими функциями, в том числе смертоносное автономное оружие и менее смертоносное оружие. Они находят применение в правоохранительных органах, а также в военной сфере и остаются практически бесконтрольными. Эти системы могут выбирать, обнаруживать, идентифицировать и атаковать цели без участия человека. Автономное оружие срабатывает благодаря датчикам и программному обеспечению, которое воспроизводит действия человека с «заданным профилем», определенным алгоритмом системы. Автономные системы вооружений имеют очень серьезные последствия для прав человека, в том числе для права на жизнь, запрета пыток и других видов жестокого обращения, а также права на личную неприкосновенность⁵⁵.

38. На заседании Первого комитета Генеральной Ассамблеи было отмечено, что в условиях, когда мир готовится к «технологическому прорыву», стремительно сокращается время для принятия мер по защите от опасностей, связанных с применением автономного оружия и искусственного интеллекта в военных целях⁵⁶. Специальный докладчик по вопросу о внесудебных казнях, казнях без надлежащего судебного разбирательства или произвольных казнях ранее рекомендовал Совету по правам человека призвать все государства объявить и соблюдать национальные моратории, по крайней мере, на испытания, производство, сборку, передачу, приобретение, принятие на вооружение и применение смертоносных автономных роботов⁵⁷.

39. Существует высокий риск серьезной и в некоторых обстоятельствах смертельной расовой дискриминации в результате использования автономных систем вооружения. Критерии, используемые для выбора целей, вероятно, включают пол, возраст и расу⁵⁸. Целевые профили также включают, казалось бы, нейтральные критерии, такие как вес или тепловые сигнатуры, однако машины часто отражают предубеждения своих программистов и общества. Они также могут быть запрограммированы на намеренно дискриминационные целевые профили⁵⁹. Например, Израиль, как сообщается, использует смертоносные автономные и полуавтономные системы вооружений. К ним предположительно относятся дистанционно управляемые квадрокоптеры для нанесения ударов по палестинцам, а также автоматизированные системы генерирования целей, с беспрецедентной скоростью составляющие объемные списки целей для поражения⁶⁰. Сообщается, что использование двух систем искусственного интеллекта Gospel и Lavender Армией обороны Израиля позволило увеличить масштабы разрушений в секторе Газа, что привело к значительным жертвам, в частности среди палестинских женщин и детей⁶¹.

⁵⁵ Amnesty International, “Autonomous weapons systems: five key human rights issues for consideration” (April 2015), p. 5.

⁵⁶ United Nations, “Without adequate guardrails, artificial intelligence threatens global security in evolution from algorithms to armaments, speaker tells First Committee”, 24 October 2023.

⁵⁷ A/HRC/23/47, п. 113.

⁵⁸ Ray Acheson, “Gender and bias”, available at <https://www.stopkillerrobots.org/wp-content/uploads/2021/09/Gender-and-Bias.pdf>.

⁵⁹ Bonnie Docherty, “Expert Panel on the Social and Humanitarian Impact of Autonomous Weapons at the Latin American and Caribbean Conference on Autonomous Weapons”, Human Rights Watch, 8 March 2023.

⁶⁰ Marwa Fatafta and Daniel Leufer, “Artificial genocidal intelligence: how Israel is automating human rights abuses and war crimes”, Access Now, 9 May 2024.

⁶¹ Yuval Abraham, “‘Lavender’: the AI machine directing Israel’s bombing spree in Gaza”, +972 Magazine, 3 April 2024.

2. Охрана здоровья

а) Рейтинг рисков для здоровья

40. Искусственный интеллект может быть использован для построения рейтинга рисков для здоровья для различных целей в системе здравоохранения, включая медицинскую диагностику и планирование ухода. Из-за особенностей алгоритмов и данных, используемых для обучения систем искусственного интеллекта, применение таких алгоритмов для распределения ресурсов в системе здравоохранения может привести к усилению расовой дискриминации. В отдельных случаях такие алгоритмы разрабатывались для определения тех, кто может претендовать на дополнительный уход, исходя из предыдущих расходов на медицинское обслуживание как косвенного показателя медицинских потребностей. На данных, лежащих в основе таких решений, может отразиться отсутствие адекватного доступа к медицинскому обслуживанию среди представителей маргинализированных расовых и этнических групп в контексте системного расизма, а также различия в структуре заболеваемости, обусловленные неравенством социально-экономических предпосылок здоровья.

41. В США был разработан калькулятор, помогающий медицинским работникам оценить вероятность успешных вагинальных родов после кесарева сечения. Исследование, проведенное в 2019 году, выявило предвзятость в базовом алгоритме калькулятора. В калькуляторе использовались два поправочных коэффициента, основанные на расовой принадлежности, которые занижали прогнозируемые показатели успешных вагинальных родов для женщин африканского происхождения и латиноамериканок по сравнению с белыми женщинами с аналогичными характеристиками. Из-за этих поправочных коэффициентов калькулятор потенциально усиливает расовые различия в сфере охраны материнского здоровья, не рекомендуя врачам предлагать вагинальные роды женщинам африканского происхождения и латиноамериканкам, что приводит к повышению частоты кесаревых сечений⁶².

б) Диагностика заболеваний

42. Еще одна область применения технологий искусственного интеллекта — диагностика заболеваний, в том числе рака⁶³. Системы искусственного интеллекта, обученные на обширных массивах данных, состоящих из тысяч или миллионов изображений, включая рентгенологические снимки, изображения патологий и фотографии, могут научиться различать доброкачественные и раковые поражения⁶⁴. Такое применение искусственного интеллекта может существенно помочь в ранней диагностике рака, потенциально спасая жизни людей и повышая эффективность системы здравоохранения. Однако представители маргинализированных расовых и этнических групп могут не воспользоваться в равной мере результатами таких достижений, поскольку алгоритмы не могут быть распространены на популяции пациентов, которые недостаточно представлены в обучающих данных. Исследователи указывают на то, что алгоритмы искусственного интеллекта для выявления рака кожи хуже работают с людьми с темным цветом кожи, поскольку многие общедоступные наборы изображений, используемые для их обучения, необъективны и не учитывают все разнообразие оттенков кожи и этнического происхождения⁶⁵. Например, исследование 21 массива данных о поражениях кожи, имеющегося в открытом доступе и содержащего более 100 000 изображений, выявило значительную недопредставленность изображений с темной кожей⁶⁶.

⁶² Darshali A. Vyas and others, “Challenging the use of race in the Vaginal Birth after Cesarean Section Calculator”, *Women’s Health Issues*, vol. 29, No. 3 (2019).

⁶³ Материалы, представленные Privacy International.

⁶⁴ Likhitha Kolla and Ravi B. Parikh, “Uses and limitations of artificial intelligence for oncology”, *Cancer*, 30 March 2024.

⁶⁵ David Wen and others, “Characteristics of publicly available skin cancer image datasets: a systematic review”, *The Lancet Digital Health*, vol. 4, No. 1 (2022).

⁶⁶ Там же. См. также материалы, представленные Privacy International.

с) Медицинские устройства с искусственным интеллектом

43. Развитие и использование искусственного интеллекта сопровождают другие достижения в области медицинских технологий, включая медицинские приборы. Поскольку многие из этих приборов оснащены искусственным интеллектом, расовая предвзятость может быть заложена в их работу. Например, в Соединенном Королевстве был опубликован доклад, свидетельствующий о том, что предвзятость закладывается в работу медицинских устройств на всех этапах их разработки, в том числе на этапах, связанных с разработкой алгоритмов и машинным обучением. Во время пандемии коронавирусной инфекции (COVID-19) использование приборов пульсоксиметрии для контроля низкого уровня кислорода в крови привело к завышению данных о содержании кислорода в крови людей с более темным цветом кожи⁶⁷.

3. Образование

а) Алгоритмы академического и карьерного успеха

44. В таких странах, как Финляндия и США, в сфере образования используются инструменты предиктивной аналитики, позволяющие определить вероятность будущих успехов на основе данных, статистических алгоритмов и машинного обучения⁶⁸. В этих алгоритмах используются данные о посещаемости, успеваемости, поведении и онлайн-активности. Они призваны помочь педагогам направлять учащихся в принятии решений, определяющих их образовательный и карьерный путь. Хотя инструменты предиктивной аналитики призваны помочь педагогам в повышении успеваемости учащихся, они часто оценивают расовые меньшинства как менее успешные в учебе и карьере из-за особенностей алгоритма и выбора данных. Опираясь на эти оценки, преподаватели могут уводить учащихся из маргинализированных расовых и этнических групп от такого образовательного и карьерного выбора, которые максимально раскрыл бы их потенциал и предоставил наилучшие возможности для преодоления циклов отчуждения, или вкладывать меньше ресурсов в этих учащихся.

б) Алгоритмы оценки успеваемости

45. Алгоритмы оценки успеваемости обычно используют исторические данные для оценки успеваемости учащихся. Такие данные могут быть необъективными из-за исторически сложившегося системного расизма в учебных заведениях. Необъективность данных будет воспроизводиться алгоритмами оценки успеваемости учащихся, особенно если исключить влияние преподавателя⁶⁹. Алгоритмы оценки успеваемости могут иметь огромное значение для определения возможностей, доступных учащимся, в том числе в отношении доступа к университетскому образованию или возможностям трудоустройства после окончания обучения. Таким образом, автоматическое принятие решений в условиях расовой предвзятости может ограничить возможности учащихся из маргинализированных расовых и этнических групп и подорвать потенциал образования как инструмента борьбы с системным расизмом.

46. Соединенное Королевство представляет собой поучительный пример применения алгоритма оценки успеваемости. В 2020 году экзамены продвинутого уровня (уровень А) были отменены из-за пандемии COVID-19. Вместо экзаменационных оценок преподавателей просили предсказывать результаты учащихся. Затем национальное агентство оценки успеваемости задействовало

⁶⁷ Материалы, представленные Privacy International.

⁶⁸ Stina Westman and others, "Artificial intelligence for career guidance – current requirements and prospects for the future", *International Academic Forum Journal of Education*, vol. 9, No. 4 (2021); и Kelli A. Bird, Benjamin L. Castleman and Yifeng Song, "Are algorithms biased in education? Exploring racial bias in predicting community college student success", *Journal of Policy Analysis and Management*, 31 January 2024.

⁶⁹ Benjamin Herold, "Why schools need to talk about racial bias in AI-powered technologies", *Education Week*, 12 April 2022.

алгоритм для стандартизации прогнозируемых оценок на основе исторических данных оценок каждой школы. У 40 % учащихся, многие из которых посещали школы в районах с низким уровнем дохода, в результате были снижены баллы. И наоборот, алгоритм повысил оценки успеваемости непропорционально большого числа учащихся из независимых платных школ. Правительство отреагировало на эту ситуацию, отменив стандартизированные результаты применения алгоритма. Однако этот эпизод привел к серьезным сбоям в процессах набора в университеты⁷⁰.

c) Большие языковые модели в сфере образования

47. Инструменты генеративного искусственного интеллекта опираются на большие языковые модели для создания нового контента, включая текст, музыку, изображения и видео. Большие языковые модели используются в образовательных учреждениях и могут помочь улучшить успеваемость учащихся всех возрастов. Исследования показали, что языковые модели небеспристрастны к английскому языку, который является самым распространенным языком в Интернете и на котором работает большинство исследователей и технологов в области искусственного интеллекта. Более того, лишь немногие из примерно 6000 языков мира могут похвастаться наличием ресурсов высококачественных данных, которые можно использовать для обучения моделей искусственного интеллекта. Чтобы устранить этот пробел, компании начали разрабатывать многоязычные языковые модели. Однако многоязычные модели работают не так хорошо, как англоязычные. Использование больших языковых моделей в образовательных учреждениях может поставить в невыгодное положение учащихся из языковой среды, не представленной в базовых ресурсах данных, что может иметь непропорциональные с расовой точки зрения последствия⁷¹.

48. Ведутся споры о том, не лучше ли запретить учащимся использовать инструменты генеративного искусственного интеллекта, основанные на больших языковых моделях, вместо того, чтобы включать их в учебные программы. В некоторых учебных заведениях также предпринимаются шаги по ограничению использования учащимися инструментов генеративного искусственного интеллекта, основанных на больших языковых моделях. Некоторые учебные заведения используют инструменты искусственного интеллекта для выявления фактов использования искусственного интеллекта учащимися. Использование таких инструментов, которые с точки зрения используемых алгоритмов могут быть предвзятыми, для выявления недобросовестных учащихся может сделать получаемые результаты еще менее объективными и навредить учащимся из маргинализированных расовых и этнических групп. Такой вред будет еще большим, если в образовательных учреждениях не предусмотрены справедливые процедуры обжалования⁷².

d) Распознавание лиц в образовательных учреждениях

49. Технологии распознавания лиц были внедрены во многих образовательных учреждениях по всему миру, несмотря на свидетельства того, что их использование не

⁷⁰ Bryan Walsh, “How an AI grading system ignited a national controversy in the U.K.”, Axios, 19 August 2020; и Daan Kolkman, “F**k the algorithm? What the world can learn from the UK’s A-level grading fiasco”, London School of Economics Impact Blog, 26 August 2020.

⁷¹ Felix Richter, “The most spoken languages: on the Internet and in real life”, Statista, 21 February 2024; Emily M. Bender, “The #BenderRule: on naming the languages we study and why it matters”, The Gradient, 14 September 2019; Gabriel Nicholas and Aliya Bhatia, “Lost in translation: large language models in non-English content analysis”, Center for Democracy and Technology, 23 May 2023; A. Bergman and Mona Diab, “Towards responsible natural language annotation for the varieties of Arabic”, in

The 60th Annual Meeting of the Association for Computational Linguistics: Findings of ACL 2022 (Association for Computational Linguistics, 2022); и BigScience Workshop, “A 176B-parameter open-access multilingual language model” (ArXiv, 2022).

⁷² См. Regina Ta and Darrell M. West, “Should schools ban or integrate generative AI in the classroom?”, Brookings Institution, 7 August 2023; и Robert Topinka, “The software says my student cheated using AI. They say they’re innocent. Who do I believe?”, *The Guardian*, 13 February 2024.

гарантирует расовой нейтральности, о чем говорилось выше. Системы распознавания лиц используются для автоматизации учета посещаемости, повышения безопасности школ, проведения экзаменов и даже для регистрации эмоций детей в школах, чтобы следить за тем, как они учатся. Зачастую это происходит без должного контроля уважения прав человека или надзора со стороны регулирующих органов. Например, в Бразилии все больше школ внедряют средства распознавания лиц для оптимизации работы, отслеживания посещаемости и повышения безопасности⁷³. Однако сообщается, что ни муниципалитеты, ни штаты не проводили исследований для оценки последствий с точки зрения соблюдения прав человека и не анализировали риски дискриминации, связанные с использованием программного обеспечения для распознавания лиц, перед реализацией этих проектов⁷⁴.

50. Использование программного обеспечения для распознавания лиц в образовательных учреждениях чревато дискриминационными в расовом плане последствиями. Известны случаи, один из которых произошел в Королевстве Нидерландов, когда студентам африканского происхождения приходилось освещать свои лица фонариком, чтобы системы искусственного интеллекта узнали их и допустили до сдачи важных экзаменов. Такой опыт негативно влияет на равноправие учащихся в сфере образования, а также становится причиной трений и отчуждения, когда у учащихся из маргинализированных расовых и этнических групп создается впечатление, что система проектировалась без учета их потребностей. Регистрация и мониторинг детских эмоций в школах серьезно влияет на право учащихся на неприкосновенность их личной жизни и может увековечить расовую предвзятость. Было установлено, что эти системы по-разному интерпретируют выражения лица людей африканского происхождения и белых людей, чаще приписывая негативные эмоции, такие как презрение и гнев, лицам африканского происхождения⁷⁵.

С. Новые инициативы по регулированию искусственного интеллекта и управлению им

51. Государства начали предпринимать многообещающие шаги по регулированию искусственного интеллекта и управлению им. В настоящем разделе Специальный докладчик обращает внимание на некоторые из этих инициатив. В основу ее анализа, который не является исчерпывающим, легли материалы, представленные государствами и группами гражданского общества, а также на ее работе в странах и исследования, проведенные для настоящего доклада.

1. Национальные инициативы

52. Государства предпринимают шаги по регулированию искусственного интеллекта и управлению им на национальном уровне, используя как обязательные правовые нормы, так и добровольные политические стандарты, а во многих случаях и их сочетание. Например, законодательные предложения о регулировании технологического пространства, включая искусственный интеллект, внесены на рассмотрение в Бразилии⁷⁶. Правительство также приняло ряд соответствующих программных документов, таких как документ под названием «Расизм в Интернете: доказательства для разработки цифровой политики», который, как сообщается, был подготовлен Министерством по вопросам расового равенства и предусматривает меры по устранению предвзятости алгоритмов, в том числе расовой⁷⁷. Специальный докладчик приветствует усилия по разработке конкретных и обязательных к исполнению нормативных положений, дополняемых соответствующими политическими стандартами. Вместе с тем она получила тревожную информацию об отсутствии эффективных консультаций с лицами африканского происхождения и их

⁷³ Материалы, представленные InternetLab.

⁷⁴ Там же.

⁷⁵ Там же.

⁷⁶ Материалы, представленные Бразилией.

⁷⁷ Материалы, представленные группой экспертов из Бразилии.

участия в разработке законодательных положений о регулировании искусственного интеллекта, а также об отсутствии общей согласованности и последовательности различных стандартов и текущей практики государства⁷⁸.

53. Другой пример — Соединенные Штаты, которые, как сообщается, предпринимают шаги для разработки комплекса обязательных и добровольных стандартов использования искусственного интеллекта. После недавнего посещения страны Специальный докладчик приветствовала подписание Указа 14110 о разработке и использовании безопасного, надежного и заслуживающего доверия искусственного интеллекта и включение в него упоминаний рисков предвзятости и дискриминации при использовании искусственного интеллекта. В ходе подготовки настоящего доклада Специальный докладчик получила дополнительную информацию о регулировании искусственного интеллекта в Соединенных Штатах, в том числе о работе над проектом закона о правах в связи с искусственным интеллектом в качестве добровольного стандарта, об усилиях по регулированию искусственного интеллекта в таких штатах, как Вирджиния, Калифорния и Нью-Йорк, и об инициативах по стимулированию коммерческих организаций к принятию на себя добровольных обязательств по разработке безопасного, надежного и прозрачного искусственного интеллекта⁷⁹. Специальный докладчик приветствует эти усилия, хотя и обеспокоена тем, что, несмотря на вызывающие доверие исследования, посвященные изучению серьезной расовой предвзятости алгоритмов в цифровых коммерческих продуктах в контексте Соединенных Штатов, в Указе № 14110 отсутствует прямая ссылка на расовую дискриминацию и предвзятость⁸⁰.

54. Сообщается, что Канада занимается разработкой обязательных стандартов в сочетании с добровольными. Закон об искусственном интеллекте и данных в настоящее время находится в стадии разработки и, как сообщается, будет включать в себя обязательные правовые меры контроля систем искусственного интеллекта с высокой степенью риска. Кроме того, Канада разработала добровольные стандарты, в том числе Добровольный кодекс поведения по ответственной разработке передовых генеративных систем искусственного интеллекта и управлению ими. Она также разработала инструмент под названием «Оценка влияния алгоритмов», который призван помочь правительственным департаментам и агентствам в оценке и смягчении рисков искусственного интеллекта, в том числе связанных с дискриминацией и предвзятостью⁸¹.

55. Специальный докладчик получила информацию о других государствах, таких как Австралия, Индия, Китай и Япония, которые, как сообщается, принимают меры для управления искусственным интеллектом и его регулированию, в том числе с помощью политических мер и, в некоторых случаях, имеющего обязательную силу законодательства⁸².

2. Региональные инициативы

56. Что касается региональных инициатив, Специальный докладчик приветствует информацию о принятии Закона об искусственном интеллекте, полученную ей от Европейского союза и его государств-членов⁸³. Она признает, что Закон является

⁷⁸ Там же.

⁷⁹ Материалы, представленные NetMission.Asia. См. также Kay Firth-Butterfield, Karen Silverman and Benjamin Larsen, “Understanding the US ‘AI Bill of Rights’ – and how it can help keep AI Accountable”, World Economic Forum, 14 October 2022; United States, Office of Science and Technology Policy of the White House, “Blueprint for an AI bill of rights: making automated systems work for the American people”, white paper, October 2022; и United States, White House, “Fact sheet: Biden-Harris Administration secures voluntary commitments from eight additional artificial intelligence companies to manage the risks posed by AI”, 12 September 2023.

⁸⁰ A/HRC/56/68/Add.1, п. 54.

⁸¹ Canada, Innovation, Science and Economic Development Canada, “The Artificial Intelligence and Data Act (AIDA) – companion document”, 13 March 2023; и материалы, представленные NetMission.Asia.

⁸² Материалы, представленные NetMission.Asia.

⁸³ Материалы, представленные Европейским союзом и Испанией.

регуляторным стандартом, который окажет значительное влияние на регион Европейского союза благодаря приведению национальных правовых норм в соответствие с его положениями. Специальный докладчик приветствует тот факт, что в тексте Закона об искусственном интеллекте учтен фактор расовой принадлежности, предусмотрены гарантии соблюдения прав человека при использовании искусственного интеллекта с высокой степенью риска, запрещены некоторые виды использования искусственного интеллекта и предусмотрены механизмы правовой защиты лиц, пострадавших от использования систем искусственного интеллекта с высокой степенью риска. Она также приветствует тот факт, что в плане действий Европейского союза по борьбе с расизмом на 2020–2025 годы говорится о расовой дискриминации, возникающей в результате использования новых технологий, таких как искусственный интеллект, что свидетельствует об определенной согласованности политики в рамках различных стандартов Европейского союза⁸⁴. В то же время Специальный докладчик получила вызывающую глубокую обеспокоенность информацию о том, что в контексте иммиграционного и пограничного контроля и правоохранительной деятельности существуют исключения из гарантий, предусмотренных Законом⁸⁵. Сообщается, что такие исключения существуют, несмотря на богатую историю расовой дискриминации в обеих этих областях и неизбежные трудности, связанные с формированием параллельных правовых систем⁸⁶. Такой подход чреват закреплением существующей расовой иерархии и сохранением практики нарушений прав человека в контексте иммиграции, пограничного контроля и правоприменения в государствах — членах Европейского союза.

3. Международные инициативы

57. Специальному докладчику известно о мерах, принимаемых Организацией Объединенных Наций для содействия управлению искусственным интеллектом. Она приветствует создание Генеральным секретарем Консультативного органа высокого уровня по искусственному интеллекту и публикацию его недавнего промежуточного доклада. Однако она сожалеет, что в этом докладе конкретно не говорится о риске расовой предвзятости и дискриминации. Специальный докладчик приветствует работу Управления Верховного комиссара Организации Объединенных Наций по правам человека (УВКПЧ), направленную на интеграцию темы прав человека в международный диалог о новых технологиях, таких как искусственный интеллект, в том числе в рамках проекта В-Tech. Помимо работы Организации Объединенных Наций, Специальному докладчику известно и о других международных инициативах, посвященных поощрению диалога и вопросам управления, включая инициативы Организации экономического сотрудничества и развития и Группы семи⁸⁷.

58. Международные организации имеют все возможности для стимулирования международного сотрудничества, технической помощи и исследований, с тем чтобы управление искусственным интеллектом осуществлялось таким образом, чтобы не усугублять и без того серьезное неравенство между странами, которое во многих случаях является наследием колониализма и рабства. Серьезные различия в технологической инфраструктуре могут привести к тому, что страны будут сталкиваться с разными проблемами при внедрении инструментов искусственного интеллекта. То внимание, которое уделяется технологиям искусственного интеллекта и проблемам дискриминации, которые они создают, в значительной степени сосредоточено на странах глобального Севера, что может привести к пробелам в понимании того, как искусственный интеллект будет влиять на культурные, религиозные и другие меньшинства глобального Юга⁸⁸. Существует риск, что

⁸⁴ Материалы, представленные Европейским союзом. См. также https://www.europarl.europa.eu/doceo/document/TA-9-2021-0290_EN.html.

⁸⁵ Материалы, представленные Privacy International; и Access Now, “The EU AI Act: a failure for human rights, a victory for industry and law enforcement”, 13 March 2024.

⁸⁶ См. A/HRC/48/76.

⁸⁷ Материалы, представленные NetMission.Asia.

⁸⁸ Danni Yu, Hannah Rosenfeld and Abhishek Gupta, “The ‘AI divide’ between the Global North and the Global South”, World Economic Forum, 16 January 2023.

наиболее развитые страны глобального Севера смогут влиять на дебаты и диалог по вопросам искусственного интеллекта таким образом, что это увековечит глобальный дисбаланс сил и ограничит возможности стран глобального Юга получать от него потенциальные выгоды.

D. Международно-правовая основа в области прав человека

59. Технологии искусственного интеллекта должны опираться на международные стандарты в области прав человека. Самый полный запрет расовой дискриминации содержится в Международной конвенции о ликвидации всех форм расовой дискриминации. Как говорится в пункте 1 статьи 1, разрабатывая Конвенцию, государства-участники стремились дать широкое определение расовой дискриминации, означающей любое различие, исключение, ограничение или предпочтение, основанное на признаках расы, цвета кожи, родового, национального или этнического происхождения, имеющие целью или следствием уничтожение или умаление признания, использования или осуществления на равных началах прав человека и основных свобод в политической, экономической, социальной, культурной или любых других областях общественной жизни.

60. Государства — участники Международной конвенции о ликвидации всех форм расовой дискриминации обязались добиваться создания внутри страны и за ее пределами сообщества, свободного от всех форм расизма. Согласно статье 2 Конвенции, государства-участники в своем стремлении содействовать реальному достижению расового равенства должны обеспечить, чтобы они не принимали участия ни в каких актах расовой дискриминации, ни в последующих программах, ведущих к расовому неравенству. Кроме того, там, где существует расизм, расовое неравенство или расовая дискриминация, государства-участники обязаны принимать эффективные и незамедлительные меры. Это обязательство действовать является абсолютным. Обязательства государств-участников предотвращать расовое неравенство и расовую дискриминацию включают в себя как превентивные, так и коррекционные действия. Хотя Конвенция обеспечивает наиболее полный запрет расовой дискриминации, другие договоры также обеспечивают защиту от таких форм дискриминации.

61. Обязательства по достижению расового равенства и обеспечению недискриминации распространяются на все сферы государственной политики и влияния, включая разработку и применение технологий искусственного интеллекта. Является ли расовая дискриминация в результате использования искусственного интеллекта преднамеренной или нет, не имеет значения для обязанности государств-участников действовать, учитывая охват запрета расовой дискриминации в соответствии с Международной конвенцией о ликвидации всех форм расовой дискриминации и другими договорами по правам человека. Обязанности государств — участников Конвенции стремиться к созданию внутрисредового и международного сообщества, свободного от всех форм расовой дискриминации, также имеют значение с точки зрения того, как государства предотвращают и устраняют неравенство внутри стран и между ними в отношении распределения плодов технологий искусственного интеллекта.

62. Государства также должны обеспечить, чтобы все расовые и этнические группы пользовались правами человека в полном объеме, как это предусмотрено в статье 5 Международной конвенции о ликвидации всех форм расовой дискриминации. Статья 5 предусматривает равенство перед законом, особенно в части права на равенство перед судом и всеми другими органами, отправляющими правосудие; право на личную безопасность и защиту со стороны государства от насилия или телесных повреждений, причиняемых как правительственными должностными лицами, так и какими бы то ни было отдельными лицами, группами или учреждениями; право на свободу мирных собраний и ассоциаций; право на здравоохранение, медицинскую помощь, социальное обеспечение и социальное обслуживание; право на образование и профессиональную подготовку. Эти права, а также положения, гарантирующие их недискриминационное применение, предусмотрены и в других договорах по правам

человека, включая Международный пакт о гражданских и политических правах и Международный пакт об экономических, социальных и культурных правах.

63. Существуют и другие положения международного права прав человека, которые возлагают на государства ответственность за устранение дискриминационных последствий применения технологий искусственного интеллекта, как говорилось выше. Сбор и использование данных без гарантий соблюдения прав человека вызывает серьезные опасения по поводу неприкосновенности частной жизни, которые особенно сильны в случае представителей маргинализированных расовых и этнических групп. Соответственно, Специальный докладчик хотела бы напомнить государствам о положениях статьи 17 Международного пакта о гражданских и политических правах, которые предусматривают запрет произвольного или незаконного вмешательства в частную жизнь человека и налагают на государства обязательство обеспечивать соответствующую правовую защиту. Другие положения Пакта также применимы к проявлениям расовой дискриминации, связанным с технологиями искусственного интеллекта. Использование искусственного интеллекта, в том числе в таких контекстах, как правоохранительная деятельность, может повлиять на свободу и безопасность личности и иметь жизненно важные последствия для представителей маргинализированных расовых и этнических групп. Статья 6 Пакта определяет неотъемлемое право на жизнь и обязывает государства обеспечивать его правовую защиту. Статья 7 закрепляет право не подвергаться пыткам и другим жестоким, бесчеловечным или унижающим достоинство видам обращения или наказания. В статье 9 говорится, что каждый человек имеет право на свободу и личную неприкосновенность и что никто не может быть подвергнут произвольному аресту или содержанию под стражей. Статья 14 гарантирует всеобщее равенство перед судами и трибуналами. Статья 26 предусматривает защиту от дискриминации групп меньшинств. Пункт 1 статьи 2 Пакта устанавливает обязательство обеспечить недискриминационное применение всех его положений. Существуют также положения международного права прав человека, касающиеся использования искусственного интеллекта в сфере иммиграционного и пограничного контроля, а также в контексте социальных сетей. Они рассматриваются в предыдущих докладах, подготовленных на основании данного мандата⁸⁹.

64. Международное право прав человека предусматривает, что все люди, которые могут подвергнуться расовой дискриминации, имеют право на доступ к средствам правовой защиты, что применимо и в случаях, когда дискриминация связана с использованием искусственного интеллекта. Статья 6 Международной конвенции о ликвидации всех форм расовой дискриминации предусматривает право на доступ к эффективной защите и средствам правовой защиты через компетентные национальные суды и другие государственные учреждения. Кроме того, Генеральная Ассамблея признала пять основных элементов права на средства правовой защиты и возмещение ущерба для жертв грубых нарушений прав человека: реституцию, компенсацию, реабилитацию, сатисфакцию и гарантии неповторения случившегося⁹⁰.

65. Коммерческие структуры играют важную роль в разработке и применении искусственного интеллекта. Они являются основными участниками процесса его разработки и часто заключают контракты с правительствами на его внедрение в государственном секторе. В Руководящих принципах предпринимательской деятельности в аспекте прав человека изложены соответствующие обязательства правительств и соответствующие обязанности в области прав человека как правительств, так и предприятий. Руководящие принципы устанавливают, что государства должны защищать от нарушений прав человека, совершаемых третьими сторонами на их территории и/или под их юрисдикцией, в том числе предприятиями. Государства должны обеспечивать такую защиту, в частности, путем проведения

⁸⁹ A/75/590 и A/78/538.

⁹⁰ Основные принципы и руководящие положения Организации Объединенных Наций, касающиеся права на правовую защиту и возмещение ущерба для жертв грубых нарушений международных норм в области прав человека и серьезных нарушений международного гуманитарного права, пп. 15–23.

эффективной политики, принятия законов, нормативных актов и вынесения судебных решений. Руководящие принципы предусматривают ответственность компаний предотвращать, смягчать и устранять нарушения прав человека, которые они могут вызвать или которым они могут способствовать, а также обеспечивать должную заботу о правах человека применительно к соответствующим видам коммерческой деятельности⁹¹. Кроме того, Руководящие принципы устанавливают обязательства правительств и ответственность бизнеса за обеспечение доступа к средствам правовой защиты в случае нарушений прав человека, связанных с предпринимательской деятельностью, дополняя право на средства правовой защиты, предусмотренное в других стандартах, о которых говорилось выше. В рамках проекта УВКПЧ B-Tech ведется работа по подготовке рекомендаций и мобилизации ресурсов для реализации Руководящих принципов в технологическом пространстве, включая конкретную работу по искусственному интеллекту⁹².

IV. Выводы и рекомендации

66. Предыдущий мандатарий обратилась к государствам и другим заинтересованным сторонам, включая коммерческие структуры, с четким призывом отказаться от «безразличного к цвету кожи» подхода к управлению развивающимися технологиями, включая искусственный интеллект, и их регулированию. Она настоятельно призвала государства регулировать эти технологии в рамках подхода, признающего структурный расизм и основанного на ключевых правозащитных стандартах. Тем не менее управление искусственным интеллектом и его регулирование во многом остаются недостаточными, неадекватно учитывающими расовую предвзятость и не отражающими международные правозащитные стандарты. Эта реальность сохраняется, несмотря на ясность и своевременность предыдущих призывов к использованию «безразличного к цвету кожи» подхода и повышение осведомленности о существовании проблемы системного расизма за прошедшие четыре года. Предположение о том, что технологии объективны и нейтральны, по-прежнему широко распространено и стимулирует погоню за внедрением искусственного интеллекта в общество, несмотря на его расово-дискриминационное воздействие и без должного рассмотрения вопроса о том, действительно ли это необходимо. Несмотря на то, что искусственный интеллект обладает положительным потенциалом, в том числе для обеспечения равенства и инклюзивности, он не является панацеей от всех социальных проблем, и им необходимо эффективно управлять, чтобы находить баланс между преимуществами и рисками.

67. Эффективное и всеобъемлющее регулирование искусственного интеллекта играет центральную роль в достижении этого тонкого баланса. Хотя эффективное регулирование искусственного интеллекта имеет жизненно важное значение, существуют дополнительные шаги, которые государства и другие стороны могут предпринять для эффективного устранения расово-дискриминационного воздействия этих технологий. Также очень важно расширять среди населения информационно-просветительскую работу по вопросам новых технологий с акцентом на права человека и повышать уровень грамотности в области искусственного интеллекта. Когда отдельные люди и группы понимают, что такое искусственный интеллект, и знают о своих правах человека в цифровом пространстве, они имеют возможность ответственно использовать эти знания и становятся разборчивой и ответственной аудиторией,

⁹¹ См. также United Nations Office on Genocide Prevention and the Responsibility to Protect and Economic and Social Research Council Human Rights, Big Data and Technology Project, University of Essex, “Countering and addressing online hate speech: a guide for policy makers and practitioners”, policy paper, July 2023; и A/74/486, пп. 44 и 45.

⁹² См. OHCHR, “B-Tech Project: OHCHR and business and human rights”, available at <https://www.ohchr.org/en/business-and-human-rights/b-tech-project>.

которая может улучшить подотчетность в сфере использования систем искусственного интеллекта.

68. Государствам следует:

a) в срочном порядке решить проблему регулирования искусственного интеллекта, учитывая скорость развития этих технологий и множество каналов, по которым они уже сегодня увековечивают расовую дискриминацию во всех сферах жизни общества;

b) разработать регуляторную базу в области искусственного интеллекта, основанную на всестороннем понимании системного расизма и опирающуюся на международное право прав человека, в том числе на запрет расовой дискриминации. Такая база должна формироваться не на основе разрозненных подходов, а с учетом различных правовых инструментов, включая специальное законодательство, посвященное искусственному интеллекту, законы о неприкосновенности частной жизни, положения о свободе информации, антидискриминационное законодательство и отраслевые нормы, чтобы добиться всеобъемлющего и эффективного регулирования, предотвращающего и устраняющего расово-дискриминационное воздействие искусственного интеллекта;

c) рассмотреть роль, которую могут сыграть в системе регулирования искусственного интеллекта добровольные стандарты. Добровольные стандарты могут служить практическим руководством по применению правовых стандартов в реальной жизни. Однако регулирование искусственного интеллекта не должно опираться только на добровольные стандарты, поскольку эти технологии имеют серьезные последствия для прав человека, в том числе в связи с расовой дискриминацией;

d) закрепить в регуляторной базе юридическое обязательство проводить комплексные оценки «должной заботы» о правах человека, включая четкие критерии оценки расовой и этнической предвзятости, при разработке и внедрении всех технологий искусственного интеллекта. Должная забота о правах человека должна включать протоколы проверки данных и пороговые значения, которые гарантировали бы защиту от алгоритмической предвзятости, в том числе расовой и этнической. Проведение этих оценок должно завершаться до внедрения новых технологий, особенно в общественном контексте, например в образовательных учреждениях, правоохранительных органах и медицинских учреждениях;

e) рассмотреть возможность запрета использования систем искусственного интеллекта, которые известны тем, что несут неприемлемые риски для прав человека, включая те, которые нарушают запрет расовой дискриминации;

f) обеспечить наличие в регуляторной базе положений, гарантирующих полную прозрачность процессов автоматизированного принятия решений, включая права на доступ к информации, в случаях, когда использование искусственного интеллекта признается допустимым по итогам комплексной оценки в рамках «должной заботы» о правах человека;

g) предусмотреть на практике четкие и доступные процедуры подачи апелляций для оценки и устранения расово-дискриминационного воздействия искусственного интеллекта, которые проводились бы с участием человека. Необходимо обеспечить равный доступ к таким процедурам обжалования;

h) создать механизмы, позволяющие пострадавшим лицам и группам получить доступ к средствам правовой защиты, обеспечивающим реституцию, компенсацию, реабилитацию, сатисфакцию и гарантии неповторения в случаях, когда технологии искусственного интеллекта привели к нарушениям прав человека, в том числе к расовой дискриминации;

i) избегать любых исключений в регуляторных стандартах, которые могут привести к нарушению запрета расовой дискриминации, предусмотренного международным правом прав человека;

j) обеспечить проведение содержательных и эффективных консультаций с заинтересованными сторонами из всех маргинализированных расовых и этнических групп, а также со специалистами в соответствующих секторах при разработке и внедрении норм регулирования искусственного интеллекта, а также при разработке и использовании технологий искусственного интеллекта;

k) инвестировать в сбор дезагрегированных данных во всех соответствующих секторах, чтобы получить информацию, необходимую для регулирования искусственного интеллекта на основе понимания системного расизма, для решения проблем с данными в системах искусственного интеллекта и для более эффективного мониторинга и оценки воздействия технологий искусственного интеллекта на представителей маргинализированных расовых и этнических групп;

l) взять на вооружение подход к данным, основанный на правозащитных стандартах, обеспечив дезагрегирование, самоидентификацию, прозрачность, конфиденциальность, участие и подотчетность при сборе и хранении данных⁹³;

m) создать надежные механизмы надзора и постоянного мониторинга инструментов искусственного интеллекта, включая регулярный аудит их воздействия, для обеспечения соблюдения норм регулирования и решения любых проблем, возникающих у затрагиваемых лиц или сообществ, а также для исправления потенциальных перекосов, создаваемых моделями искусственного интеллекта с течением времени;

n) участвовать в международном сотрудничестве, укреплении потенциала и проведении исследований, чтобы обеспечить более справедливое распределение плодов искусственного интеллекта между странами и избежать углубления неравенства, которое существует как наследие колониализма и рабства;

o) развивать среди общественности имеющую правозащитный акцент просветительскую работу, посвященную приемлемому и ответственному использованию технологий искусственного интеллекта, для повышения грамотности в области искусственного интеллекта, которая должна включать компоненты, конкретно направленные на повышение осведомленности о расово-дискриминационном воздействии искусственного интеллекта.

69. Коммерческим организациям следует:

a) проводить оценку «должной заботы» о соблюдении прав человека на всех этапах проектирования, разработки и внедрения продуктов искусственного интеллекта;

b) обеспечить содержательные и эффективные консультации с представителями маргинализированных расовых и этнических групп, специалистами из соответствующих областей общества и теми, кто обладает опытом борьбы с системным расизмом, при проектировании, разработке и внедрении продуктов искусственного интеллекта;

c) разработать протоколы для обеспечения полной прозрачности и обмена информацией об алгоритмическом принятии решений в отношении продуктов, которые имеют последствия для прав человека;

d) обеспечить постоянный мониторинг продуктов искусственного интеллекта на предмет расовой предвзятости;

⁹³ См. A/HRC/42/59 и A/HRC/44/57.

e) разработать программу подготовки по проблемам расовой дискриминации, включая неявные предубеждения и системный расизм, для всех, кто участвует в проектировании, разработке и внедрении искусственного интеллекта. При разработке таких программ подготовки следует опираться как на международные стандарты в области прав человека, так и на результаты исследований, посвященных расово-дискриминационному воздействию технологий искусственного интеллекта;

f) содействовать усилиям по развитию имеющих правозащитную направленность программ просвещения общественности с целью расширения возможностей для повышения грамотности в области искусственного интеллекта.

70. Организации Объединенных Наций и ее независимым правозащитным механизмам следует:

a) содействовать эффективному диалогу и дискуссиям среди заинтересованных сторон, посвященным технологиям искусственного интеллекта и их регулированию;

b) привести в работу Консультативного органа высокого уровня по искусственному интеллекту четкий акцент на расово-дискриминационные последствия применения технологий искусственного интеллекта;

c) обеспечить, чтобы публикации и руководства по технологиям искусственного интеллекта основывались на принципах международного права прав человека, в том числе на запрете расовой дискриминации, и прямо признавали, что сохранение расовых перекосов при разработке и внедрении технологий искусственного интеллекта является серьезной глобальной проблемой;

d) играть роль в мониторинге последствий применения технологий искусственного интеллекта для прав человека, в том числе связанных с расовой дискриминацией;

e) поддерживать международное сотрудничество, укрепление потенциала и исследовательскую работу, чтобы попытаться добиться более справедливого распределения плодов искусственного интеллекта между странами.