

**Consejo de Derechos Humanos****56º período de sesiones**

18 de junio a 14 de julio de 2024

Tema 9 de la agenda

**Racismo, discriminación racial, xenofobia y formas
conexas de intolerancia: seguimiento y aplicación
de la Declaración y el Programa de Acción de Durban****Formas contemporáneas de racismo, discriminación racial,
xenofobia y formas conexas de intolerancia****Informe de la Relatora Especial sobre las formas contemporáneas
de racismo, discriminación racial, xenofobia y formas conexas
de intolerancia, Ashwini K. P.****Resumen*

En el presente informe, la Relatora Especial sobre las formas contemporáneas de racismo, discriminación racial, xenofobia y formas conexas de intolerancia, Ashwini K. P., resume las actividades que ha llevado a cabo durante el último año y aborda las formas en que la suposición predominante de que la tecnología es objetiva y neutra está propiciando que la inteligencia artificial perpetúe la discriminación racial. La Relatora Especial examina cuatro formas transversales en las que la inteligencia artificial puede contribuir a las manifestaciones de discriminación racial: los problemas relacionados con los datos, las cuestiones de diseño de algoritmos, el uso intencionadamente discriminatorio de la inteligencia artificial y las cuestiones relativas a la rendición de cuentas. A continuación, ofrece ejemplos de la aplicación de la inteligencia artificial en diversos ámbitos sociales y de sus repercusiones discriminatorias desde el punto de vista racial. Asimismo, analiza los esfuerzos emergentes para gestionar y regular la inteligencia artificial antes de ofrecer una visión general de las normas pertinentes del derecho internacional de los derechos humanos. La Relatora Especial concluye presentando recomendaciones sobre cómo deben enfocar los Estados la gestión y regulación de las tecnologías de inteligencia artificial para prevenir y abordar la discriminación racial.

* Se acordó publicar este documento tras la fecha prevista debido a circunstancias que escapan al control de quien lo presenta.



I. Introducción

1. Este informe se presenta en cumplimiento de la resolución 52/36 del Consejo de Derechos Humanos, en la que el Consejo solicitó a la Relatoría Especial sobre las formas contemporáneas de racismo, discriminación racial, xenofobia y formas conexas de intolerancia que le presentara un informe anual. En el informe, la Relatora Especial describe las actividades que ha llevado a cabo en el marco de su mandato y explora el tema de la inteligencia artificial y la discriminación racial.

2. Para elaborar el informe, la Relatora Especial invitó a los Estados Miembros de las Naciones Unidas y otras partes interesadas, incluidas organizaciones de la sociedad civil, organizaciones internacionales e instituciones nacionales de derechos humanos, a aportar comunicaciones. La Relatora Especial expresa su sincera gratitud a todos los Estados Miembros y demás partes interesadas que presentaron información. La Relatora Especial tuvo en cuenta dichas contribuciones para elaborar el informe y sigue abierta a un diálogo permanente con todas las partes interesadas sobre esta importante cuestión¹.

II. Resumen de las actividades

3. En octubre de 2023, la Relatora Especial presentó a la Asamblea General, en su septuagésimo octavo período de sesiones, sus informes sobre la lucha contra la glorificación del nazismo, el neonazismo y otras prácticas que contribuyen a exacerbar las formas contemporáneas de racismo, discriminación racial, xenofobia y formas conexas de intolerancia y sobre el discurso de odio racista en línea². Entre el 31 de octubre y el 14 de noviembre de 2023, la Relatora Especial realizó una visita a los Estados Unidos de América³.

4. En agosto de 2023, la Relatora Especial participó en la novena sesión del Grupo de Eminentes Expertos Independientes sobre la Aplicación de la Declaración y el Programa de Acción de Durban. En enero de 2024, la Relatora Especial asistió a la reunión regional para Asia y el Pacífico sobre el Decenio Internacional de los Afrodescendientes. En febrero de 2024, asistió a la Conferencia Internacional sobre Justicia Alimentaria desde la Perspectiva de los Derechos Humanos, celebrada en Qatar y centrada en el tema “Desafíos de la realidad y apuestas futuras”. En abril de 2024, asistió al tercer período de sesiones del Foro Permanente sobre los Afrodescendientes, donde presentó una ponencia sobre la superación del racismo sistémico y los daños históricos en la educación.

III. Inteligencia artificial y discriminación racial

5. En el presente informe, la Relatora Especial ha decidido centrarse en la inteligencia artificial y la discriminación racial. Este tema está en consonancia con su enfoque estratégico sobre el nexo entre las tecnologías digitales y la discriminación racial, como se indica en el informe que presentó al Consejo de Derechos Humanos en su 53º período de sesiones, en el que expuso su visión estratégica y sus prioridades iniciales⁴. La Relatora Especial aprovecha la labor de la anterior titular del mandato relativa a las nuevas tecnologías digitales y la

¹ La Relatora Especial recibió apoyo en materia de investigación y análisis por parte de la Clínica Internacional de Derechos Humanos de la Facultad de Derecho de Harvard y de la Clínica Internacional de Derechos Humanos y Resolución de Conflictos y el Centro Stanford para la Justicia Racial de la Facultad de Derecho de Stanford. Agradece sinceramente a todos los participantes su inestimable apoyo en la preparación del informe.

² [A/78/302](#) y [A/78/538](#).

³ Véase [A/HRC/56/68/Add.1](#).

⁴ [A/HRC/53/60](#), párrs. 50 a 53.

discriminación racial⁵ y responde al interés del Consejo de Derechos Humanos y del sistema de las Naciones Unidas en general respecto a la gobernanza de la inteligencia artificial⁶.

6. Las novedades en el campo de la inteligencia artificial generativa y su floreciente aplicación siguen planteando graves problemas de derechos humanos, incluida la preocupación por la discriminación racial. La inteligencia artificial generativa está cambiando el mundo y tiene el potencial de impulsar cambios sociales cada vez más sísmicos en el futuro. La rápida difusión de la aplicación de la inteligencia artificial en diversos ámbitos preocupa profundamente a la Relatora Especial. Esto no se debe a que la inteligencia artificial carezca de beneficios potenciales. De hecho, presenta posibles oportunidades de innovación e inclusión. Sin embargo, estas tecnologías están creciendo y evolucionando con un ímpetu en gran medida desenfrenado. A la Relatora Especial le preocupa que las medidas normativas y jurídicas que tratan de gestionar y regular la inteligencia artificial no avanza al mismo ritmo que el crecimiento de esta tecnología y que los nuevos esfuerzos por gobernarla y regularla no presten suficiente atención a su enorme capacidad actual y a su potencial futuro para perpetuar y profundizar la discriminación racial sistémica, así como para ampliar la desigualdad dentro de las regiones, los países y las comunidades y entre ellos.

7. Tal y como expuso la anterior titular del mandato, existe una noción tan persistente como perjudicial de que la tecnología es neutra y objetiva:

La percepción pública suele ser que la tecnología es intrínsecamente neutra y objetiva; algunos han señalado que esta presunción de objetividad y neutralidad tecnológica es una característica que aún sigue siendo una de las más destacables incluso entre los creadores de tecnologías. Ahora bien, la tecnología nunca es neutra, pues refleja los valores e intereses de quienes influyen en su diseño y uso, y está fundamentalmente conformada por las mismas estructuras de desigualdad que operan en la sociedad.⁷

8. En el presente informe, la Relatora Especial aborda las formas en que la suposición predominante de que la tecnología es objetiva y neutra está propiciando que la inteligencia artificial perpetúe la discriminación racial.

A. Formas transversales en que la inteligencia artificial puede contribuir a las manifestaciones de discriminación racial

9. La inteligencia artificial no es un monolito. De hecho, hay varios tipos. La inteligencia artificial predictiva, que se considera una forma "tradicional" de esta tecnología, se basa en modelos que utilizan datos históricos, patrones y tendencias para hacer predicciones fundamentadas respecto a resultados o acontecimientos futuros.

10. La inteligencia artificial que se utiliza para identificar caracteres impresos, rostros humanos, objetos y otra información es otra forma de inteligencia artificial "tradicional" y engloba diversas tecnologías para reconocer y distinguir objetos, individuos y patrones en los datos que se le facilitan.

11. Los sistemas de inteligencia artificial generativa son formas más recientes de inteligencia artificial. Son versátiles y pueden utilizarse para diversos fines. Abarcan una clase de sistemas de inteligencia artificial diseñados para producir diversos resultados sobre la base de amplios conjuntos de datos de entrenamiento, redes neuronales, arquitectura de aprendizaje profundo e indicaciones del usuario. Los modelos de inteligencia artificial generativa pueden producir una amplia gama de resultados, como imágenes, texto, audio, vídeo y datos sintéticos. A diferencia de los modelos de inteligencia artificial que se centran en detectar patrones en los datos existentes, la inteligencia artificial generativa se entrena para crear nuevos puntos de datos que imiten los patrones y las características de los datos

⁵ Véanse [A/75/590](#), [A/HRC/44/57](#) y [A/HRC/48/76](#).

⁶ Véase, por ejemplo, Órgano Asesor de Alto Nivel sobre Inteligencia Artificial, "Interim report: governing AI for humanity" (diciembre de 2023); resolución 53/29 del Consejo de Derechos Humanos; y resoluciones de la Asamblea General 78/213 y 78/265.

⁷ [A/HRC/44/57](#), párr. 12.

utilizados para entrenar los modelos de aprendizaje automático. La llegada de la inteligencia artificial generativa dará lugar a muchas nuevas aplicaciones, así como a muchas nuevas cuestiones de derechos humanos⁸.

12. Estos distintos tipos de inteligencia artificial tienen multitud de aplicaciones. A continuación, la Relatora Especial expone ejemplos más concretos de los usos de la inteligencia artificial y las implicaciones conexas en términos de discriminación racial. No obstante, desea subrayar que es muy importante examinar los aspectos comunes de las formas en que la inteligencia artificial puede perpetuar la discriminación racial, en particular en el marco de los debates jurídicos y normativos relativos a su gestión y regulación. En esos debates, los efectos de la inteligencia artificial deben verse a través del prisma del racismo sistémico, que se define como “el funcionamiento de un sistema complejo e interrelacionado de leyes, políticas, prácticas y actitudes en las instituciones del Estado, el sector privado y las estructuras sociales que, combinadas, dan lugar a una discriminación, distinción, exclusión, restricción o preferencia, directa o indirecta, intencionada o no, de hecho o de derecho, por motivos de raza, color, ascendencia u origen nacional o étnico”⁹. Como se refleja en esa definición, el racismo sistémico es un fenómeno complejo, a menudo insidioso y que afecta a toda la sociedad. Las manifestaciones de racismo sistémico en un ámbito están interrelacionadas, son interdependientes y se refuerzan mutuamente con las de otros ámbitos. Examinar las formas transversales en que la inteligencia artificial contribuye a la discriminación racial puede ayudar a establecer las formas en que interactúa con las manifestaciones del racismo sistémico y las refuerza, y en que intensifica de forma holística la opresión sistémica en la sociedad según criterios raciales y étnicos¹⁰.

1. Problemas relacionados con los datos

13. El auge de los sistemas de inteligencia artificial y los algoritmos de aprendizaje automático ha llevado a la digitalización de datos a escala masiva. Los algoritmos utilizan esos datos para tomar decisiones y emprender acciones en diversos sectores. Sin embargo, los conjuntos de datos con los que se entrenan los algoritmos suelen ser incompletos o infrarrepresentan a determinados grupos de personas. Si determinados grupos están sobrerrepresentados o infrarrepresentados en los conjuntos de entrenamiento, incluidos los grupos raciales y étnicos, puede producirse un sesgo algorítmico. Del mismo modo, si los conjuntos de entrenamiento incluyen datos ya sesgados, pueden producir resultados sesgados.

14. Si los datos de entrenamiento son insuficientes, los algoritmos pueden hacer predicciones sistemáticamente discriminatorias para grupos no representados o infrarrepresentados en los datos. El sesgo algorítmico no solo puede producirse con muy pocos datos; los algoritmos basados en datos no representativos también pueden producir resultados sesgados. Por ejemplo, en un estudio centrado en las bases de datos de imágenes de las fuerzas del orden en los Estados Unidos se demostró que las personas afrodescendientes tenían más probabilidades de ser señaladas erróneamente en las redes de reconocimiento facial utilizadas por los agentes del orden. Esto se debió a errores en la identificación facial de ese grupo y a la sobrerrepresentación de afrodescendientes en las bases de datos de fotografías policiales, que refleja pautas históricas de racismo sistémico¹¹.

15. Los sesgos históricos pueden afectar a los propios datos. Un elemento central del aprendizaje automático es hacer predicciones sobre el futuro a partir de datos del pasado. Sin embargo, si los datos del pasado están sesgados en contra de ciertos grupos, en particular por criterios raciales y étnicos, los modelos informáticos pueden reproducir y amplificar esos sesgos. El uso de datos sesgados o erróneos para tomar decisiones en la vida real puede perjudicar aún más a los grupos raciales y étnicos marginados, porque el uso de esos datos

⁸ Comunicación de la Comisión de Derechos Humanos de Australia. Todas las comunicaciones se publicarán en el sitio web de la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (ACNUDH).

⁹ A/HRC/47/53, párr. 9.

¹⁰ A/HRC/44/57, párr. 43.

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

en el contexto de la inteligencia artificial crea más datos, que luego se utilizan para sustentar decisiones futuras. Estos sistemas que se refuerzan a sí mismos pueden reproducir y profundizar las disparidades existentes.

16. La última cuestión relacionada con los datos es la privacidad. Los datos utilizados en los sistemas de inteligencia artificial suelen incluir información personal de las personas a las que pertenecen. La recogida y el tratamiento de datos sin consentimiento vulnera el derecho a la intimidad. También se dan casos de datos recogidos en un entorno, como la asistencia sanitaria, incluso mediante el uso de aplicaciones sanitarias, que se comparten, sin consentimiento, para su uso en otros, como por ejemplo con fines policiales. Las filtraciones de datos y el acceso no autorizado a información personal a través de piratería informática plantean problemas de privacidad adicionales. Para las personas pertenecientes a grupos marginados por motivos raciales, los problemas de derechos humanos relacionados con el derecho a la intimidad pueden verse amplificadas. Las violaciones de la intimidad pueden poner a estos grupos en peligro de ostracismo, discriminación o peligro físico¹².

2. Problemas de diseño de algoritmos

17. Una segunda forma común de sesgo en las herramientas de inteligencia artificial surge de la forma en que se diseñan los algoritmos. Si el sesgo está implícito en las decisiones de diseño, un algoritmo puede contribuir a obtener resultados sesgados, aunque los datos introducidos en el algoritmo sean perfectamente representativos. Las decisiones sobre los parámetros y el funcionamiento de un algoritmo pueden introducir sesgos. Los diseñadores de algoritmos toman decisiones sobre qué variables utilizará un algoritmo, cómo definir categorías o umbrales para clasificar la información y qué datos se utilizarán para construir el algoritmo. Entre las decisiones que toman los diseñadores está cómo medir características específicas y definir el éxito algorítmico. A veces, los antecedentes o las perspectivas de los diseñadores de algoritmos pueden hacer que incorporen sesgos inconscientes, incluidos sesgos raciales, en sus diseños de algoritmos¹³. Al parecer, esta falta de diversidad en los sectores de la tecnología digital se ve agravada por la ausencia de procesos de consulta inclusivos en el desarrollo de los sistemas de inteligencia artificial, lo que contribuye a los problemas de diseño algorítmico¹⁴.

18. Las elecciones relativas al diseño algorítmico pueden tener importantes efectos discriminatorios en la vida real. Por ejemplo, al crear un algoritmo de evaluación del riesgo de préstamo, la forma de definir y medir el “riesgo” puede dar lugar a resultados discriminatorios. Si el diseñador de un algoritmo decide utilizar la puntuación crediticia como variable indirecta del riesgo, podrían darse resultados discriminatorios respecto a los grupos de personas que tienden a tener puntuaciones crediticias más bajas. Algunas investigaciones han demostrado que puede existir una fuerte correlación entre la puntuación crediticia, la raza y otros indicadores demográficos, y que el uso de puntuaciones crediticias perjudica a determinados grupos¹⁵. Esa correlación puede considerarse, en muchos casos, un subproducto del racismo y la exclusión sistémicos existentes. Las personas pueden verse perjudicadas por la decisión de un diseñador de algoritmos de utilizar puntuaciones crediticias para evaluar el riesgo de los préstamos, a pesar de que aparentemente no sea un criterio discriminatorio.

3. Utilización con fines discriminatorios

19. En algunos casos, la inteligencia artificial puede utilizarse con fines explícitamente racistas a través de su despliegue selectivo contra grupos específicos, lo que da lugar a resultados discriminatorios. Por ejemplo, hay informes de que las fuerzas del orden utilizan intencionadamente la inteligencia artificial para vigilar y sobrevigilar a determinadas

¹² Samantha Lai y Brooke Tanner, “Examining the intersection of data privacy and civil rights”, Brookings Institution, 18 de julio de 2022. Véase también la comunicación de Privacy International.

¹³ Ninareh Mehrabi y otros, “A survey on bias and fairness in machine learning”, *ACM Computing Surveys*, vol. 54, núm. 6 (2022); comunicación de The London Story; y [A/HRC/44/57](#), párr. 17.

¹⁴ Comunicación de NetMission.Asia.

¹⁵ A. R. Lange y Natasha Duarte, “Understanding bias in algorithmic design”, Medium, 6 de septiembre de 2017.

comunidades, siguiendo criterios de discriminación racial¹⁶. Además, la discriminación intencionada puede producirse cuando los gobiernos y otras entidades aprovechan las capacidades de la tecnología para vigilar o perfilar a grupos o individuos específicos, o centrarse en ellos, en función de su identidad racial o étnica¹⁷.

20. La difusión de desinformación es otra forma de utilizar la inteligencia artificial con fines explícitamente racistas. Los agentes políticos pueden utilizar la inteligencia artificial para generar textos, imágenes y vídeos con el fin de manipular la opinión pública y los procesos políticos a su favor y socavar la confianza en las instituciones, entre otros por motivos raciales. También se informa de gobiernos que han utilizado la inteligencia artificial para sembrar la discordia y facilitar la censura en línea¹⁸.

4. Problemas relativos a la rendición de cuentas

21. El hecho de que algunas herramientas de inteligencia artificial tomen decisiones independientemente de los humanos significa que el proceso de toma de decisiones está oculto, como en una “caja negra” opaca. Además, un algoritmo podría tomar decisiones de forma independiente porque, una vez expuestos a los datos, los algoritmos de inteligencia artificial se actualizan constantemente. Con el tiempo, una herramienta de inteligencia artificial puede utilizar, en su toma de decisiones, factores en los que no estaba programada originalmente para basarse. En cambio, estos factores proceden de patrones que ella misma ha detectado en los datos. A medida que el algoritmo incorpora estos nuevos patrones a su código y a su toma de decisiones, es posible que las personas que confían en él ya no puedan “mirar bajo el capó” y definir con precisión los criterios que el algoritmo ha utilizado para ofrecer determinados resultados. Así, el problema de la “caja negra” hace que el procedimiento de razonamiento de la inteligencia artificial sea insidioso y opaco¹⁹. Además, muchos algoritmos desarrollados por entidades empresariales no pueden ser examinados debido a las leyes sobre contratos y propiedad intelectual, lo que agrava los problemas de rendición de cuentas²⁰.

22. El problema de la “caja negra” tiene implicaciones especialmente preocupantes en el contexto del racismo sistémico. Como se ha descrito anteriormente, el racismo sistémico es una lacra insidiosa pero profundamente destructiva que afecta a toda la sociedad. No siempre se reconocen las fuerzas que impulsan el racismo sistémico, un fenómeno que se ve agravado por las lagunas persistentes en la recopilación de datos desglosados por raza y etnia²¹. La inteligencia artificial, sin mecanismos eficaces de rendición de cuentas, tiene una capacidad significativa para ser un motor adicional del ya insidioso y destructivo fenómeno del racismo sistémico.

23. Los problemas relacionados con la rendición de cuentas que entraña la inteligencia artificial tienen implicaciones considerables respecto a la capacidad de quienes sufren actos de discriminación racial para buscar recursos efectivos. Hoy en día, cuando las personas pertenecientes a grupos raciales y étnicos marginados experimentan diversos resultados derivados de la toma de decisiones humanas, los tribunales y otros mecanismos de rendición de cuentas pueden examinar si las acciones fueron intencionadas y justificables²². Cuando son las personas las que toman las decisiones, a menudo hay pruebas que pueden utilizarse para hacer esas evaluaciones. En muchos casos, los procesos autónomos de toma de decisiones no dejan rastros probatorios del mismo modo que lo hacen las personas

¹⁶ Véase Amnistía Internacional, *Decode Surveillance NYC: Methodology* (Londres, 2022); y comunicación de NetMission.Asia.

¹⁷ Comunicación de NetMission.Asia.

¹⁸ Tate Ryan-Mosley, “[How generative AI is boosting the spread of disinformation and propaganda](#)”, *MIT Technology Review*, 4 de octubre de 2023.

¹⁹ Yavar Bathaee, “The artificial intelligence black box and the failure of intent and causation”, *Harvard Journal of Law and Technology*, vol. 31, núm. 2 (2018); [A/HRC/44/57](#), párr. 34; y Renata M.

O'Donnell, “Challenging racist predictive policing algorithms under the Equal Protection Clause”, *New York University Law Review*, vol. 94, núm. 3 (junio de 2019).

²⁰ [A/HRC/44/57](#), párr. 44.

²¹ [A/HRC/47/53](#), párr. 16.

²² Bathaee, “The artificial intelligence black box”.

responsables de decisiones²³. Estos problemas de “caja negra” agravarán las ya importantes barreras en el acceso a la justicia para quienes sufren discriminación racial.

B. La inteligencia artificial: utilización e impacto discriminatorio

24. En esta sección, la Relatora Especial ofrece ejemplos de la aplicación de la inteligencia artificial en diversos ámbitos sociales y de sus repercusiones discriminatorias desde el punto de vista racial. Estos ejemplos son ilustrativos y no exhaustivos, y se ofrecen como prueba evidente de que la inteligencia artificial ya está contribuyendo a la discriminación racial. La Relatora Especial percibe estos ejemplos como manifestaciones de discriminación racial interconectadas y que se refuerzan mutuamente, lo que contribuye al refuerzo holístico de la opresión sistémica que afecta a toda la sociedad según criterios raciales y étnicos.

25. La Relatora Especial ha elegido tres ámbitos para ejemplificar el impacto discriminatorio de la inteligencia artificial: la aplicación de la ley, la seguridad y el sistema de justicia penal; la educación; y la atención de la salud. En relación con el uso de la inteligencia artificial en otros contextos, la Relatora Especial recomienda consultar los informes de la anterior titular del mandato sobre el aumento de las fronteras digitales y la cartografía de la discriminación racial y xenófoba en el control digital de las fronteras y de la inmigración y sobre el uso de las tecnologías digitales en la aplicación de la legislación en materia de control de fronteras e inmigración²⁴. La Relatora Especial también remite a los lectores al informe que presentó a la Asamblea General, en su septuagésimo octavo período de sesiones, sobre el discurso de odio racista en línea, en el que se aborda el uso de la inteligencia artificial en la moderación del contenido de las redes sociales²⁵, y al informe que el Relator Especial sobre la extrema pobreza y los derechos humanos presentó a la Asamblea General en su septuagésimo cuarto período de sesiones, en el que se ofrece un análisis del uso de la inteligencia artificial en los sistemas de protección social²⁶.

1. Aplicación de la ley, seguridad y sistema de justicia penal

a) Identificación automática

26. Las fuerzas del orden utilizan herramientas de identificación automatizadas para relacionar lo que observan en un entorno concreto con una posible “coincidencia” en una base de datos. Uno de los tipos más comunes de herramientas de identificación automatizada es la tecnología de reconocimiento facial. Las herramientas de reconocimiento facial toman imágenes de vídeo o fotografías de una persona y las comparten con algoritmos. A continuación, los algoritmos comparan las imágenes con una base de datos de fotografías policiales, fotografías de carnés de conducir u otras imágenes con el objetivo de identificar a la persona²⁷. Los diseñadores de estas herramientas entrenan los modelos en los que se basan mostrándoles imágenes de rostros, mediante un proceso de aprendizaje automático. El objetivo es entrenar los modelos para que identifiquen los rasgos distintivos de los rostros humanos²⁸. Sin embargo, los conjuntos de datos de imágenes utilizados para entrenar estos modelos no siempre son representativos desde el punto de vista demográfico²⁹. En un estudio de una popular base de datos de imágenes, los investigadores detectaron una sobrerrepresentación de hombres de entre 18 y 40 años y una infrarrepresentación de personas de piel oscura³⁰. Según otro estudio de los sistemas comerciales de reconocimiento

²³ *Ibid.*

²⁴ [A/75/590](#) y [A/HRC/48/76](#).

²⁵ [A/78/538](#).

²⁶ [A/74/493](#).

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

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

²⁹ Comunicación de AI for the People.

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

facial, los algoritmos de clasificación de género se entrenan con conjuntos de datos que reúnen, de manera abrumadora, rostros masculinos blancos³¹. La falta de diversidad racial, de género y cultural en los conjuntos de entrenamiento de las herramientas de inteligencia artificial conduce a uno de los clásicos problemas de datos descritos anteriormente. El algoritmo tiene más probabilidades de ofrecer coincidencias erróneas respecto a los grupos infrarrepresentados en los datos de entrenamiento, incluidos los que sufren formas interseccionales de discriminación.

27. Se ha informado de que la identificación errónea de rostros a través de estas tecnologías ha provocado un aumento del número de detenciones de afrodescendientes³². El Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión y la Alta Comisionada de las Naciones Unidas para los Derechos Humanos han señalado que las herramientas de reconocimiento facial contribuyen a menudo a la discriminación ilegal y al perfilamiento racial³³. A pesar de estas preocupaciones por los derechos humanos, las fuerzas del orden de varios países han desplegado sistemas de reconocimiento facial. Por ejemplo, el Gobierno de la India ha realizado cuantiosas inversiones en este tipo de sistemas. Según se informa, el sistema de reconocimiento facial utilizado por la policía de Delhi solo es preciso en el 2 % de los casos y expone a las comunidades minoritarias a un riesgo desproporcionado de identificación errónea y detención ilegal³⁴. Al parecer, las fuerzas del orden brasileñas han incurrido en acusaciones y detenciones ilegales basándose en herramientas de reconocimiento facial defectuosas. Según un estudio de 2019, el 90 % de las personas detenidas en ciudades brasileñas gracias a la tecnología de reconocimiento facial son afrodescendientes³⁵.

28. Los sistemas de detección de disparos son otro tipo común de herramienta de detección automatizada que utilizan agentes del orden en varios países. Un sistema, denominado ShotSpotter, consiste en colocar en los barrios sensores que contienen un micrófono, un sistema GPS, dispositivos de memoria y procesamiento y capacidad celular³⁶. Cuando los sensores detectan un ruido que podría ser un disparo, un algoritmo triangula la ubicación de lo que haya causado el ruido. El algoritmo filtra otros posibles sonidos antes de enviar el audio a una persona para que lo revise³⁷. La información disponible sugiere que los sistemas de detección de disparos se despliegan de forma desproporcionada en comunidades habitadas por grupos racialmente marginados³⁸, y que pueden tener un porcentaje de error muy elevado. La colocación de sistemas de detección de disparos en comunidades en las que viven grupos raciales y étnicos marginados y las imprecisiones de los sistemas de detección de disparos exacerban los prejuicios sistémicos de las propias fuerzas del orden.

29. Hay muchos ejemplos de cómo el uso de la tecnología de identificación automatizada ha tenido consecuencias vitales. En 2019, en los Estados Unidos, un hombre negro de Nueva Jersey fue supuestamente detenido y encarcelado ilegalmente durante diez días por un error de reconocimiento facial. A pesar de la existencia de pruebas exculpatorias, las autoridades tardaron casi un año en retirar el caso, y el afectado se enfrentaba a penas de hasta 25 años de prisión por los cargos que se le imputaban. El incidente tuvo un impacto considerable en

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

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

³³ Véanse A/HRC/41/35 y A/HRC/48/31.

³⁴ Amnistía Internacional, “Ban the scan: Hyderabad”, disponible en <https://banthescan.amnesty.org/hyderabad/>.

³⁵ Comunicación de un grupo de expertos del Brasil.

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

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

³⁸ *Ibid.*; y MacArthur Justice Center, “ShotSpotter is deployed overwhelmingly in Black and Latinx neighborhoods in Chicago”, disponible en <https://endpolicesurveillance.com/burden-on-communities-of-color/>.

su vida³⁹. Al parecer, en febrero de 2024, agentes de la ley de Chicago, en respuesta a una falsa alerta de ShotSpotter, abrieron fuego contra un niño que estaba encendiendo fuegos artificiales⁴⁰. Otro ejemplo del uso de este tipo de tecnología de inteligencia artificial es la supuesta adopción por parte de las Fuerzas de Defensa de Israel de Wolf Pack, una vasta base de datos que contiene imágenes y toda la información disponible sobre los palestinos de la Ribera Occidental, y que comprende además diversos programas de vigilancia como Blue Wolf y Red Wolf⁴¹. Al parecer, las Fuerzas de Defensa de Israel instalaron en el casco antiguo de Hebrón cámaras dotadas de inteligencia artificial capaces de identificar rostros humanos y conectadas al programa Blue Wolf, una aplicación móvil que permite a los soldados detectar y clasificar a los palestinos de toda la Ribera Occidental mediante una amplia base de datos biométricos respecto a la que la mayoría de personas no ha dado su consentimiento, lo que se traduce en una vigilancia continua de los palestinos. La rigurosa aplicación del sistema Wolf Pack por parte de las Fuerzas de Defensa de Israel agrava el *apartheid* perpetuado contra los palestinos⁴². Estos ejemplos muestran las graves implicaciones para los derechos humanos que se derivan del uso de sistemas de inteligencia artificial para tomar decisiones consecuentes en entornos de alto riesgo.

b) Algoritmos policiales predictivos

30. Otra forma de tecnología de inteligencia artificial utilizada habitualmente por las fuerzas de seguridad es la vigilancia policial predictiva. Las herramientas de predicción policial evalúan quién podría cometer delitos en el futuro y dónde podrían producirse, basándose en datos personales y de localización.

31. La vigilancia policial predictiva puede exacerbar el exceso histórico de vigilancia de comunidades según criterios raciales y étnicos⁴³. Dado que los funcionarios encargados de hacer cumplir la ley han centrado históricamente su atención en esos barrios, los miembros de las comunidades presentes en ellos están sobrerrepresentados en los registros policiales. Esto, a su vez, repercute en los lugares donde los algoritmos predicen que se producirá la delincuencia en el futuro, lo que conduce a un mayor despliegue policial en las zonas en cuestión⁴⁴. La labor policial predictiva también puede reflejar aspectos del problema de la “caja negra”, ya que los algoritmos carecen de transparencia, entre otros respecto a los datos que se analizan y al modo en que se utilizan las predicciones⁴⁵.

32. Los algoritmos de predicción policial basados en la localización se basan en vínculos entre lugares, sucesos y datos históricos sobre delincuencia para predecir cuándo y dónde es probable que se produzcan delitos en el futuro⁴⁶. Sobre esa base, las fuerzas policiales planifican sus patrullas en consecuencia. Cuando los agentes de los barrios con exceso de vigilancia registran nuevos delitos, se crea un bucle de retroalimentación, por el que el algoritmo genera predicciones cada vez más sesgadas centradas en esos barrios. En resumen, los prejuicios del pasado conducen a prejuicios en el futuro. En el Reino Unido de Gran Bretaña e Irlanda del Norte, mediante un estudio encargado por el Gobierno sobre el sesgo algorítmico en la actuación policial se demostró que la categorización de lugares geográficos

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

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

⁴¹ Amnistía Internacional, *Automated Apartheid: How Facial Recognition Fragments, Segregates and Controls Palestinians in the OPT* (Londres, 2023), págs. 41 a 45.

⁴² Sophia Goodfriend, “Algorithmic State violence: automated surveillance and Palestinian dispossession in Hebron's Old City”, *International Journal of Middle East Studies*, vol. 55, núm. 3 (2023).

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

⁴⁴ Kristian Lum y William Isaac, “To predict and serve?”, *Significance*, vol. 13, núm. 5 (2016); y comunicación de la Comisión de Derechos Humanos de Australia.

⁴⁵ Lau, “Predictive policing explained”.

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

como “zona crítica” de la delincuencia podía inducir a los agentes a esperar que se produjeran más delitos en esas zonas. En consecuencia, era más probable que los agentes detuvieran o arrestaran a personas en las “zonas críticas” por prejuicios que por auténticos imperativos de seguridad pública⁴⁷. Investigadores del Uruguay han descubierto que los datos utilizados en los algoritmos policiales predictivos basados en la localización podrían estar sesgados. La variable de ubicación podría funcionar como variable indirecta del origen socioeconómico o étnico, lo que desembocaría en resultados discriminatorios⁴⁸.

33. Las herramientas policiales predictivas basadas en datos personales buscan predecir quién podría cometer un delito en el futuro a partir de los datos básicos de los individuos. Esos datos pueden incluir la edad, el sexo, el estado civil, el historial de abuso de sustancias y los antecedentes penales de una persona. Al igual que ocurre con las herramientas basadas en la localización, los datos sobre detenciones anteriores, que a menudo están contaminados por el racismo sistémico del sistema de justicia penal, pueden sesgar las predicciones futuras de esos algoritmos. Factores como el origen socioeconómico, el nivel educativo y la ubicación pueden actuar como variables indirectas de la raza y perpetuar sesgos históricos⁴⁹. En Australia, la Policía de Nueva Gales del Sur utilizó el Plan de Gestión de Objetivos Sospechosos, basado en algoritmos, para identificar a personas susceptibles de cometer delitos. Al parecer, su uso dio lugar a un número desproporcionadamente alto de interacciones policiales con miembros de comunidades aborígenes e isleñas del estrecho de Torres antes de que dejara de utilizarse⁵⁰.

c) Algoritmos de evaluación de la reincidencia

34. Las herramientas de evaluación de la reincidencia se utilizan para fundamentar las decisiones en distintas fases del sistema de justicia penal, como la fianza, la condena y la libertad condicional⁵¹. Las herramientas de evaluación de la reincidencia utilizan datos históricos para evaluar la probabilidad de que los acusados actúen de determinadas maneras, en particular si es probable que cometan un nuevo delito en el futuro. Las herramientas generan puntuaciones de riesgo a partir de información procedente de fuentes como los antecedentes penales y las encuestas a los acusados⁵².

35. Las herramientas de predicción de la reincidencia presentan múltiples problemas de inteligencia artificial que contribuyen a la discriminación racial. En primer lugar, las herramientas muestran problemas en relación con los datos. Los datos del sistema de justicia penal utilizados para entrenar sus algoritmos reflejan desigualdades sistémicas basadas en un historial de comportamientos policiales racistas⁵³. Además, las opciones de diseño, como la forma en que se miden o evalúan las variables, pueden contribuir a la discriminación algorítmica⁵⁴. Asimismo, la forma en que el diseñador de un algoritmo decide definir el “éxito” puede influir en lo que el algoritmo trata de maximizar. Si un algoritmo está configurado para optimizar un número mínimo de nuevos delitos, puede correlacionar condenas más largas con menores tasas de reincidencia, porque las personas no pueden reincidir mientras están encarceladas. Si es así, podrá utilizar esos patrones para recomendar condenas más largas.

⁴⁷ *Ibid.* Véase también Gobierno del Reino Unido de Gran Bretaña e Irlanda del Norte, “Report commissioned by CDEI calls for measures to address bias in police use of data analytics”, 16 de septiembre de 2019.

⁴⁸ Juan Ortiz Freuler y Carlos Iglesias, “Algoritmos e inteligencia artificial en Latinoamérica: Un estudio de implementaciones por parte de Gobiernos de Argentina y Uruguay”, World Wide Web Foundation, septiembre de 2018; y Eticas Foundation, “El Ministerio del Interior de Uruguay invierte en policía predictiva”, 13 de septiembre de 2021.

⁴⁹ Heaven, “Predictive policing algorithms are racist”.

⁵⁰ Comunicación de la Comisión de Derechos Humanos de Australia.

⁵¹ Julia Angwin y otros, “Machine bias”, ProPublica, 23 de mayo de 2016.

⁵² *Ibid.*

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

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

36. Algunos investigadores han sugerido que los predictores de reincidencia no son precisos y que sus errores tienen un impacto desproporcionado en los grupos racialmente marginados. Por ejemplo, un estudio realizado en los Estados Unidos reveló que las puntuaciones de riesgo eran muy poco fiables para pronosticar delitos violentos. Se clasificaba erróneamente a los afrodescendientes como futuros delincuentes con casi el doble de frecuencia que a los blancos.

d) Sistemas de armas autónomos

37. Los sistemas de armas autónomos incluyen todo sistema de armas con autonomía en sus funciones críticas, incluidas las armas autónomas letales y las armas menos letales. Tienen aplicaciones en contextos policiales, así como militares, y permanecen en gran medida sin control. Estos sistemas pueden seleccionar, detectar, identificar y atacar objetivos sin intervención humana. Un arma autónoma se activa mediante sensores y programas informáticos que determinan la coincidencia de una persona con un "perfil de objetivo" establecido por el algoritmo del sistema. Los sistemas de armamento autónomos tienen implicaciones muy graves para los derechos humanos, entre ellas las relacionadas con el derecho a la vida, la prohibición de la tortura y otros malos tratos y el derecho a la seguridad de la persona⁵⁵.

38. Un orador relató a la Primera Comisión de la Asamblea General que la ventana de oportunidad para promulgar salvaguardias contra los peligros de las armas autónomas y las aplicaciones militares de la inteligencia artificial se estaba cerrando rápidamente, a medida que el mundo se preparaba para una "ruptura tecnológica"⁵⁶. El Relator Especial sobre las ejecuciones extrajudiciales, sumarias o arbitrarias ya recomendó en el pasado al Consejo de Derechos Humanos que instara a todos los Estados a que decretaran y aplicaran moratorias nacionales sobre, al menos, el ensayo, la producción, el ensamblaje, la transferencia, la adquisición, el despliegue y el uso de robótica autónoma letal⁵⁷.

39. Existe un grave riesgo de discriminación racial grave y, en algunas circunstancias, mortal, derivado del uso de sistemas de armas autónomos. Es muy probable que los criterios utilizados para seleccionar los objetivos incluyan el sexo, la edad y la raza⁵⁸. Los perfiles de objetivos también incluyen criterios aparentemente neutros, como el peso o el perfil térmico, pero las máquinas suelen reflejar los prejuicios de sus programadores y de la sociedad. También pueden programarse con perfiles de objetivos intencionadamente discriminatorios⁵⁹. Por ejemplo, Israel utiliza sistemas de armas letales autónomos y semiautónomos. Según se informa, esto incluye el uso de cuadricópteros teledirigidos contra objetivos palestinos, además de sistemas automatizados de generación de objetivos, que operan a una velocidad y volumen sin precedentes, para producir "listas de asesinatos"⁶⁰. Según se informa, Gospel y Lavender, dos sistemas tecnológicos de inteligencia artificial utilizados por las Fuerzas de Defensa de Israel, han intensificado los niveles de destrucción en Gaza y provocado numerosas bajas, en particular entre las mujeres y los niños palestinos⁶¹.

⁵⁵ Amnistía Internacional, "Autonomous weapons systems: five key human rights issues for consideration" (abril de 2015), pág. 5.

⁵⁶ Naciones Unidas, "Without adequate guardrails, artificial intelligence threatens global security in evolution from algorithms to armaments, speaker tells First Committee", 24 de octubre de 2023.

⁵⁷ A/HRC/23/47, párr. 113.

⁵⁸ Ray Acheson, "Gender and bias", disponible en <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 de marzo de 2023.

⁶⁰ Marwa Fatafta y Daniel Leufer, "Artificial genocidal intelligence: how Israel is automating human rights abuses and war crimes", Access Now, 9 de mayo de 2024.

⁶¹ Yuval Abraham, "'Lavender': the AI machine directing Israel's bombing spree in Gaza", +972 Magazine, 3 de abril de 2024.

2. Atención de la salud

a) Puntuaciones de riesgo sanitario

40. La inteligencia artificial puede utilizarse con miras a establecer puntuaciones de riesgo para la salud con diversos fines sanitarios, como el diagnóstico médico y la planificación de los cuidados. Cuando estos algoritmos se utilizan para asignar recursos sanitarios pueden producirse efectos de discriminación racial, debido al diseño algorítmico y a los datos utilizados para entrenar los sistemas de inteligencia artificial. Hay casos en los que se han diseñado algoritmos de este tipo para determinar quién debe tener derecho a cuidados extraordinarios, utilizando los gastos sanitarios anteriores como variable indirecta de las necesidades médicas. Los datos en los que se basan tales decisiones pueden verse influidos por la anterior falta de acceso adecuado a la atención de la salud entre las personas de grupos raciales y étnicos marginados en el contexto del racismo sistémico, así como por los diferentes patrones de enfermedad debidos a las disparidades en los determinantes socioeconómicos de la salud.

41. En los Estados Unidos se desarrolló una calculadora para ayudar a los proveedores de atención de la salud a evaluar la probabilidad de éxito de un parto vaginal tras cesárea. Un estudio de 2019 reveló sesgos en el algoritmo de base de la calculadora. La calculadora tenía dos factores de corrección basados en la raza, que dieron como resultado una predicción de tasas de éxito de parto vaginal más bajas para las mujeres afrodescendientes e hispanas en comparación con las mujeres blancas de características similares. Debido a estos factores de corrección, la calculadora empeoró potencialmente las disparidades raciales en los resultados de salud materna al disuadir a los médicos de ofrecer un parto vaginal a las mujeres afrodescendientes e hispanas, lo que llevó a tasas más altas de cesáreas⁶².

b) Detección de enfermedades

42. Otra aplicación de las tecnologías de inteligencia artificial es la detección de enfermedades, entre ellas el cáncer⁶³. Los sistemas de inteligencia artificial entrenados con amplios conjuntos de datos que comprenden miles o millones de imágenes, incluidas exploraciones radiológicas, imágenes patológicas y fotografías, pueden aprender a distinguir entre lesiones normales y cancerosas⁶⁴. Este despliegue de inteligencia artificial puede contribuir significativamente a la detección precoz del cáncer, lo que podría salvar vidas y mejorar la eficiencia del sistema de salud. Sin embargo, es posible que las personas pertenecientes a grupos raciales y étnicos marginados no se beneficien en igual medida de estos avances debido a que los algoritmos no son extensibles a poblaciones de pacientes que no están adecuadamente representadas en los datos de entrenamiento. Algunos investigadores han sugerido que el uso de algoritmos de inteligencia artificial para la detección del cáncer de piel muestra peores resultados en personas con tonos de piel más oscuros porque muchos de los conjuntos de datos de imágenes disponibles públicamente que se utilizan para entrenarlos están sesgados, con falta de diversidad en cuanto a tonos de piel y orígenes étnicos⁶⁵. Por ejemplo, un estudio de 21 conjuntos de datos de libre acceso sobre lesiones cutáneas, que contenían más de 100.000 imágenes, reveló una importante infrarrepresentación de imágenes de pieles más oscuras⁶⁶.

c) Dispositivos médicos con inteligencia artificial

43. La inteligencia artificial se está desarrollando y utilizando junto con otros avances en tecnologías de asistencia médica, incluidos los dispositivos de atención de la salud. Muchos de estos dispositivos están dotados de inteligencia artificial, y los prejuicios raciales pueden

⁶² Darshali A. Vyas y otros, "Challenging the use of race in the Vaginal Birth after Cesarean Section Calculator", *Women's Health Issues*, vol. 29, núm. 3 (2019).

⁶³ Comunicación de Privacy International.

⁶⁴ Likhitha Kolla y Ravi B. Parikh, "Uses and limitations of artificial intelligence for oncology", *Cancer*, 30 de marzo de 2024.

⁶⁵ David Wen y otros, "Characteristics of publicly available skin cancer image datasets: a systematic review", *The Lancet Digital Health*, vol. 4, núm. 1 (2022).

⁶⁶ *Ibid.* Véase también la comunicación de Privacy International.

influir en su funcionamiento. Por ejemplo, en el Reino Unido, un informe demostró que había un sesgo presente en el funcionamiento de los dispositivos médicos en todas las fases de su desarrollo, incluso durante las fases que implicaban el desarrollo de algoritmos y el aprendizaje automático. Durante la pandemia de enfermedad por coronavirus (COVID-19), el uso de dispositivos de pulsioximetría para medir los niveles bajos de oxígeno en la sangre llevó a sobrestimar los niveles de oxígeno en la sangre de las personas con tonos de piel más oscuros⁶⁷.

3. Educación

a) Algoritmos de éxito académico y profesional

44. En países como los Estados Unidos y Finlandia se utilizan herramientas de análisis predictivo en la educación para determinar la probabilidad de éxito futuro a partir de datos, algoritmos estadísticos y aprendizaje automático⁶⁸. Los datos utilizados en estos algoritmos incluyen datos sobre asistencia, calificaciones, comportamiento y actividad en línea. Están diseñados para ayudar a los educadores a orientar a los estudiantes en las decisiones sobre su trayectoria educativa y profesional. Si bien las herramientas de análisis predictivo pretenden ayudar a los educadores a mejorar los resultados de los estudiantes, a menudo califican a las minorías raciales como menos propensas a tener éxito académico y profesional, debido al diseño de los algoritmos y a la elección de los datos. Sobre la base de estas calificaciones, los educadores pueden desviar a los estudiantes correspondientes a grupos raciales y étnicos marginados de las opciones educativas y profesionales que maximizarían su potencial y ofrecerían las mejores oportunidades para romper los ciclos de exclusión, o invertir menos recursos en estos estudiantes.

b) Algoritmos de calificación

45. Los algoritmos de calificación suelen utilizar datos históricos de calificaciones para evaluar el rendimiento de los alumnos. Estos datos pueden estar sesgados por patrones históricos de racismo sistémico en las instituciones educativas. El sesgo de los datos se reproducirá en los algoritmos de puntuación predictiva de los alumnos, especialmente cuando se excluya la aportación del profesor⁶⁹. Los algoritmos de calificación pueden tener enormes consecuencias a la hora de determinar las oportunidades disponibles para los estudiantes, incluso en relación con el acceso a la educación universitaria o las oportunidades de empleo tras la educación. Por lo tanto, las decisiones automatizadas con sesgo racial pueden limitar las oportunidades de los estudiantes de grupos raciales y étnicos marginados y socavar el potencial de la educación como herramienta para acabar con el racismo sistémico.

46. El Reino Unido ofrece un ejemplo que puede servir de advertencia respecto al despliegue de un algoritmo de clasificación. En 2020, los exámenes de nivel avanzado (A-level) se cancelaron debido a la pandemia de COVID-19. En lugar de las notas de los exámenes, se pidió a los profesores que predijeran los resultados de los alumnos. A continuación, la agencia nacional reguladora de las calificaciones recurrió a un algoritmo para ajustar las puntuaciones previstas a partir de los datos históricos de calificaciones de cada centro. El 40 % de los estudiantes, muchos de los cuales asistían a escuelas situadas en zonas de renta baja, vieron rebajadas sus puntuaciones como consecuencia de ello. Por el contrario, el algoritmo otorgó una mayor nota a un número desproporcionadamente alto de estudiantes de colegios independientes de pago. El Gobierno respondió a la polémica

⁶⁷ Comunicación de Privacy International.

⁶⁸ Stina Westman y otros, "Artificial intelligence for career guidance - current requirements and prospects for the future", *International Academic Forum Journal of Education*, vol. 9, núm. 4 (2021); y Kelli A. Bird, Benjamin L. Castleman y Yifeng Song, "Are algorithms biased in education? Exploring racial bias in predicting community college student success", *Journal of Policy Analysis and Management*, 31 de enero de 2024.

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

anulando el ajuste realizado mediante el algoritmo. Sin embargo, el episodio causó importantes trastornos en los procesos de admisión a las universidades⁷⁰.

c) Grandes modelos lingüísticos y educación

47. Las herramientas de inteligencia artificial generativa se basan en los grandes modelos lingüísticos para producir contenidos novedosos, como texto, música, imágenes y vídeos. Los grandes modelos lingüísticos se utilizan en entornos educativos y pueden ayudar a mejorar los resultados académicos de alumnos de todas las edades. Los estudios han demostrado que los modelos lingüísticos están sesgados hacia el inglés, que es la lengua más utilizada en Internet y en la que trabajan la mayoría de los investigadores y tecnólogos de inteligencia artificial. Además, solo un puñado de las aproximadamente 6.000 lenguas del mundo dispone de recursos de datos de alta calidad que puedan utilizarse para entrenar modelos de inteligencia artificial. Para colmar esa laguna, las empresas han empezado a desarrollar modelos lingüísticos multilingües. Sin embargo, los modelos multilingües no funcionan tan bien como los modelos en inglés. El uso de grandes modelos lingüísticos en entornos educativos podría perjudicar a los estudiantes de entornos lingüísticos que no están representados en los recursos de datos subyacentes, lo que podría tener repercusiones desproporcionadas desde el punto de vista racial⁷¹.

48. Existen debates sobre si las herramientas de inteligencia artificial generativa basadas en grandes modelos lingüísticos deberían prohibirse entre los estudiantes en lugar de integrarse en los planes de estudios. En algunos entornos educativos también se están tomando medidas para intentar restringir entre los estudiantes el uso de herramientas de inteligencia artificial generativa que se basan en grandes modelos lingüísticos. Algunas instituciones educativas están utilizando herramientas de inteligencia artificial para detectar el uso de inteligencia artificial por parte de los alumnos. El uso de este tipo de herramientas, que pueden contener sesgos algorítmicos, para vigilar las trampas puede introducir más sesgos que perjudiquen a los estudiantes de grupos raciales y étnicos marginados. Este perjuicio se agravará en los casos en que las instituciones no hayan establecido procedimientos de recurso equitativos⁷².

d) Reconocimiento facial en centros educativos

49. Las tecnologías de reconocimiento facial se han implantado en muchos entornos educativos de todo el mundo, a pesar de haberse demostrado el sesgo racial con el que operan, descrito anteriormente. Los sistemas de reconocimiento facial se están utilizando para automatizar el registro de la asistencia, mejorar la seguridad en los colegios, supervisar exámenes e incluso registrar las emociones de los niños en las escuelas para controlar cuánto están aprendiendo. A menudo, todo ello se lleva a cabo sin la debida diligencia en materia de derechos humanos ni una supervisión reglamentaria adecuada. En el Brasil, por ejemplo, cada vez más colegios adoptan herramientas de reconocimiento facial para agilizar las operaciones, controlar la asistencia y mejorar la seguridad⁷³. Sin embargo, se ha denunciado que ni los municipios ni los estados realizaron estudios de evaluación del impacto en los

⁷⁰ Bryan Walsh, “How an AI grading system ignited a national controversy in the U.K.”, *Axios*, 19 de agosto de 2020; y 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 de agosto de 2020.

⁷¹ Felix Richter, “The most spoken languages: on the Internet and in real life”, *Statista*, 21 de febrero de 2024; Emily M. Bender, “The #BenderRule: on naming the languages we study and why it matters”, *The Gradient*, 14 de septiembre de 2019; Gabriel Nicholas y Aliya Bhatia, “Lost in translation: large language models in non-English content analysis”, *Center for Democracy and Technology*, 23 de mayo de 2023; A. Bergman y Mona Diab, “Towards responsible natural language annotation for the varieties of Arabic”, en *The 60th Annual Meeting of the Association for Computational Linguistics: Findings of ACL 2022* (Association for Computational Linguistics, 2022); y BigScience Workshop, “A 176B-parameter open-access multilingual language model” (ArXiv, 2022).

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

⁷³ Comunicación de Internet Lab.

derechos humanos ni analizaron los riesgos de discriminación asociados al programa informático de reconocimiento facial antes de poner en marcha estos proyectos⁷⁴.

50. El uso de programas informáticos de reconocimiento facial en entornos educativos está teniendo repercusiones discriminatorias desde el punto de vista racial. Ha habido casos, entre ellos uno denunciado en el Reino de los Países Bajos, en los que estudiantes afrodescendientes han tenido que iluminarse la cara con luces a fin de que los sistemas de inteligencia artificial utilizados para conceder el acceso a exámenes importantes pudieran reconocerlos. Tales experiencias repercuten en la igualdad de derechos de los alumnos a la educación, pero también crean fricción y exclusión cuando los alumnos de grupos raciales y étnicos marginados tienen la impresión de que el sistema no ha sido diseñado para ellos. La grabación y vigilancia de las emociones de los niños en las escuelas tiene importantes implicaciones para la intimidad de todos los alumnos y puede perpetuar los prejuicios raciales. Se ha comprobado que estos sistemas interpretan de forma diferente las expresiones faciales de individuos afrodescendientes y blancos, atribuyendo con más frecuencia sentimientos negativos, como el desprecio y la ira, a los afrodescendientes⁷⁵.

C. Nuevas iniciativas para regular y gestionar la inteligencia artificial

51. Los Estados han empezado a tomar medidas prometedoras para regular y gestionar la inteligencia artificial. En la presente sección, la Relatora Especial llama la atención sobre algunas de estas iniciativas. Su análisis, no exhaustivo, se basa en las comunicaciones de los Estados y de grupos de la sociedad civil, así como en su trabajo en los países y en las investigaciones realizadas para el presente informe.

1. Iniciativas nacionales

52. Los Estados han tomado medidas para regular y gestionar la inteligencia artificial a nivel nacional, tanto a través de disposiciones legales vinculantes como de normas políticas voluntarias, así como, en muchos casos, de una mezcla de ambas. Por ejemplo, en el Brasil hay disposiciones legislativas pendientes sobre la regulación del espacio tecnológico⁷⁶, incluida la inteligencia artificial. El Gobierno también ha adoptado una serie de documentos normativos relevantes, como el titulado “Racismo en Internet: pruebas para la formulación de políticas digitales”, que al parecer fue elaborado por el Ministerio de Igualdad Racial y contiene medidas para abordar el sesgo algorítmico, también en relación con la raza⁷⁷. La Relatora Especial acoge con satisfacción los esfuerzos para desarrollar disposiciones reglamentarias específicas y vinculantes, complementadas por las normas políticas pertinentes. Sin embargo, recibió información preocupante sobre la supuesta falta de consulta efectiva y participación de los afrodescendientes en el desarrollo de disposiciones legislativas sobre la regulación de la inteligencia artificial, así como la falta de coherencia y consistencia generales entre las diferentes normas y las prácticas actuales del Estado⁷⁸.

53. Otro ejemplo son los Estados Unidos, donde, según se informa, se han tomado medidas para elaborar una combinación de normas vinculantes y voluntarias sobre el uso de la inteligencia artificial. Tras su reciente visita al país, la Relatora Especial acogió con satisfacción la firma del Decreto núm. 14110 sobre el Desarrollo y Uso Seguro y Confiable de la Inteligencia Artificial y las referencias que en él se hacen a los riesgos de sesgo y discriminación en el uso de la inteligencia artificial. Durante la preparación del presente informe, la Relatora Especial recibió más información sobre la regulación de la inteligencia artificial en los Estados Unidos, en particular sobre los trabajos relativos a una declaración de derechos de la inteligencia artificial, como norma voluntaria, los esfuerzos realizados en estados como Virginia, California y Nueva York para regular la inteligencia artificial y las iniciativas encaminadas a facilitar las promesas de contribuciones voluntarias de las

⁷⁴ *Ibid.*

⁷⁵ *Ibid.*

⁷⁶ Comunicación del Brasil.

⁷⁷ Comunicación de un grupo de expertos del Brasil.

⁷⁸ *Ibid.*

entidades empresariales para desarrollar una inteligencia artificial segura y transparente⁷⁹. La Relatora Especial acoge con satisfacción estos esfuerzos, aunque le preocupa que, a pesar de la sólida investigación sobre el profundo sesgo racial algorítmico de los productos comerciales digitales en el contexto de los Estados Unidos, no haya una referencia explícita a la discriminación racial y el sesgo en el Decreto núm. 14110⁸⁰.

54. Según parece, el Canadá está elaborando una combinación de normas vinculantes y voluntarias. De acuerdo con la información disponible, la Ley de Inteligencia Artificial y Datos, que se encuentra actualmente en fase de borrador, incluirá controles legales vinculantes para los sistemas de inteligencia artificial de alto riesgo. Además, el Canadá ha elaborado normas voluntarias, entre ellas el Código de Conducta Voluntario sobre el Desarrollo y la Gestión Responsables de los Sistemas Avanzados de Inteligencia Artificial Generativa. También ha puesto en marcha la Evaluación del Impacto Algorítmico, una herramienta diseñada para ayudar a los departamentos y agencias gubernamentales a evaluar y mitigar los riesgos de la inteligencia artificial, incluidos los relacionados con la discriminación y el sesgo⁸¹.

55. La Relatora Especial recibió información sobre otros Estados, como Australia, China, la India y el Japón, que, al parecer, han adoptado medidas para gestionar y regular la inteligencia artificial, incluso mediante políticas y, en algunos casos, legislación vinculante⁸².

2. Iniciativas regionales

56. En cuanto a las iniciativas regionales, la Relatora Especial acoge con satisfacción la información que recibió de la Unión Europea, así como de sus Estados miembros, sobre la aprobación de la Ley de Inteligencia Artificial⁸³. Reconoce que la Ley es una norma reglamentaria vinculante que tendrá un impacto significativo en la región de la Unión Europea a través de la armonización de las normas jurídicas nacionales con sus disposiciones. La Relatora Especial celebra que el texto de la Ley de Inteligencia Artificial incorpore la raza, tenga salvaguardias de derechos humanos para los usos de alto riesgo de la inteligencia artificial, prohíba algunos usos de la inteligencia artificial y prevea mecanismos de recurso para las personas afectadas por el uso de sistemas de inteligencia artificial de alto riesgo. También celebra que, según se informa, el plan de acción contra el racismo de la Unión Europea para 2020-2025 aborde la discriminación racial derivada del uso de las nuevas tecnologías, como la inteligencia artificial, lo que sugiere cierto grado de coherencia normativa entre las distintas normas de la Unión Europea⁸⁴. Sin embargo, la Relatora Especial recibió información muy preocupante que indicaba que existen excepciones a las protecciones establecidas en la Ley en los contextos de la gestión de la inmigración y las fronteras y la aplicación de la ley⁸⁵. Al parecer, estas excepciones existen a pesar de la importante discriminación racial histórica en ambos ámbitos y de los riesgos que conlleva permitir el desarrollo de marcos jurídicos paralelos⁸⁶. Con este enfoque se corre el riesgo de afianzar las jerarquías raciales existentes y de perpetuar de forma significativa las violaciones

⁷⁹ Comunicación de NetMission.Asia. Véase también Kay Firth-Butterfield, Karen Silverman y Benjamin Larsen, “Understanding the US ‘AI Bill of Rights’ - and how it can help keep AI Accountable”, Foro Económico Mundial, 14 de octubre de 2022; Estados Unidos, Oficina de Política Científica y Tecnológica de la Casa Blanca, “Blueprint for an AI bill of rights: making automated systems work for the American people”, libro blanco, octubre de 2022; y Estados Unidos, Casa Blanca, “Fact sheet: Biden-Harris Administration secures voluntary commitments from eight additional artificial intelligence companies to manage the risks posed by AI”, 12 de septiembre de 2023.

⁸⁰ [A/HRC/56/68/Add.1](#), párr. 54.

⁸¹ Canadá, Innovación, Ciencia y Desarrollo Económico del Canadá, “The Artificial Intelligence and Data Act (AIDA) - companion document”, 13 de marzo de 2023; y comunicación de NetMission.Asia.

⁸² Comunicación de NetMission.Asia.

⁸³ Comunicaciones de la Unión Europea y España.

⁸⁴ Comunicación de la Unión Europea. Véase también https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html.

⁸⁵ Comunicación de Privacy International; y Access Now, “The EU AI Act: a failure for human rights, a victory for industry and law enforcement”, 13 de marzo de 2024.

⁸⁶ Véase [A/HRC/48/76](#).

de los derechos humanos en los contextos de la gestión de la inmigración y las fronteras y de la aplicación de la ley en todos los Estados miembros de la Unión Europea.

3. Iniciativas internacionales

57. La Relatora Especial es consciente de las medidas adoptadas por las Naciones Unidas para contribuir a la gestión de la inteligencia artificial. Se congratula de que el Secretario General creara el Órgano Asesor de Alto Nivel sobre Inteligencia Artificial y de la publicación de su reciente informe provisional. Sin embargo, lamenta que en dicho informe no se haga referencia específica al riesgo de prejuicios raciales y discriminación. La Relatora Especial acoge con satisfacción la labor de la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (ACNUDH) dirigida a integrar los derechos humanos en los diálogos internacionales sobre las tecnologías emergentes, como la inteligencia artificial, en particular a través del Proyecto B-Tech. Además de la labor de las Naciones Unidas, la Relatora Especial está al corriente de otras iniciativas internacionales para promover el diálogo y la gestión, incluidas las iniciativas de la Organización de Cooperación y Desarrollo Económicos y del Grupo de los Siete⁸⁷.

58. Las organizaciones internacionales están bien posicionadas para facilitar la cooperación internacional, la asistencia técnica y la investigación destinadas a garantizar que la inteligencia artificial se gobierne de manera que no exacerbe las ya de por sí grandes desigualdades entre países, que existen en muchos casos como legado del colonialismo y la esclavitud. Las diferencias significativas en infraestructura tecnológica pueden plantear distintos retos a los países a la hora de implantar herramientas de inteligencia artificial. Gran parte de la atención prestada a la tecnología de inteligencia artificial y a los problemas de discriminación que plantea se ha centrado en los países del Norte Global, lo que podría dar lugar a lagunas en la comprensión de cómo la inteligencia artificial repercutirá en las minorías culturales, religiosas y de otro tipo en el Sur Global⁸⁸. Existe el riesgo de que los países más desarrollados del Norte Global puedan influir en el debate y el diálogo sobre la inteligencia artificial de forma que se perpetúen los desequilibrios de poder a escala mundial y se limite la capacidad de los países del Sur Global para aprovechar los beneficios potenciales.

D. Marco jurídico internacional de derechos humanos

59. La tecnología que se sustente en inteligencia artificial debe basarse en el derecho internacional de los derechos humanos. La prohibición más completa de la discriminación racial figura en la Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial. Como se refleja en el párrafo 1 de su artículo 1, los Estados redactaron la Convención de manera que incluyera una definición amplia de discriminación racial, entendida como toda distinción, exclusión, restricción o preferencia basada en motivos de raza, color, linaje u origen nacional o étnico que tenga por objeto o por resultado anular o menoscabar el reconocimiento, goce o ejercicio, en condiciones de igualdad, de los derechos humanos y libertades fundamentales en las esferas política, económica, social, cultural o en cualquier otra esfera de la vida pública.

60. Los Estados partes en la Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial se han comprometido a perseguir el logro de una comunidad nacional e internacional libre de todas las formas de racismo. Con miras a facilitar el logro sustantivo de la igualdad racial, de conformidad con el artículo 2 de la Convención, los Estados partes deben garantizar que no participarán en ningún acto de discriminación racial ni fomentarán programas que conduzcan a la desigualdad racial. Además, cuando existe racismo, desigualdad racial o discriminación racial, los Estados partes tienen la obligación de adoptar medidas eficaces e inmediatas. Esta obligación de actuar es absoluta. Las obligaciones de los Estados partes de prevenir la desigualdad racial y la discriminación racial abarcan tanto acciones preventivas como correctivas. Si bien la Convención establece

⁸⁷ Comunicación de NetMission.Asia.

⁸⁸ Danni Yu, Hannah Rosenfeld y Abhishek Gupta, "The 'AI divide' between the Global North and the Global South", Foro Económico Mundial, 16 de enero de 2023.

la prohibición más completa de la discriminación racial, otros tratados también ofrecen protección frente a tales formas de discriminación.

61. Las obligaciones de lograr la igualdad racial y garantizar la no discriminación se extienden a todos los ámbitos de la política y la influencia de los gobiernos, incluido el diseño y la aplicación de tecnologías de inteligencia artificial. El hecho de que la discriminación racial resultante de la inteligencia artificial sea intencionada o no es irrelevante en relación con el deber de actuar de los Estados partes, dado el alcance de la prohibición de la discriminación racial en virtud de la Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial y otros tratados de derechos humanos. Los deberes de los Estados partes en la Convención de procurar el logro de una comunidad nacional e internacional libre de todas las formas de discriminación racial también son pertinentes para la forma en que los Estados previenen y abordan las desigualdades dentro de los países y entre ellos en relación con la distribución de los beneficios de las tecnologías de inteligencia artificial.

62. Los Estados también deben garantizar que todos los grupos raciales y étnicos disfruten plenamente de sus derechos humanos, tal y como se recoge en el artículo 5 de la Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial. El artículo 5 contempla la igualdad ante la ley, en particular en cuanto al disfrute de los derechos a la igualdad de tratamiento en los tribunales y todos los demás órganos que administran justicia; a la seguridad personal y a la protección del Estado contra todo acto de violencia o atentado contra la integridad personal cometido por funcionarios públicos o por cualquier individuo, grupo o institución; a la libertad de reunión y de asociación pacíficas; a la salud pública, la asistencia médica, la seguridad social y los servicios sociales; y a la educación y la formación profesional. Estos derechos, así como las disposiciones que garantizan su aplicación no discriminatoria, también están contemplados en otros tratados de derechos humanos, como el Pacto Internacional de Derechos Civiles y Políticos y el Pacto Internacional de Derechos Económicos, Sociales y Culturales.

63. Existen otras disposiciones del derecho internacional de los derechos humanos que confieren a los Estados la responsabilidad de abordar los efectos discriminatorios de las tecnologías de inteligencia artificial, como se ha descrito anteriormente. La recopilación y el uso de datos sin salvaguardias de los derechos humanos plantean importantes problemas de privacidad, que pueden verse amplificadas en el caso de las personas pertenecientes a grupos raciales y étnicos marginados. En consecuencia, la Relatora Especial desea recordar a los Estados las disposiciones del artículo 17 del Pacto Internacional de Derechos Civiles y Políticos, que prevén el derecho a no ser objeto de injerencias arbitrarias o ilegales en la vida privada y que imponen a los Estados la obligación de garantizar la protección jurídica pertinente. Otras disposiciones del Pacto también se aplican a las manifestaciones de discriminación racial relacionadas con las tecnologías de inteligencia artificial. El uso de inteligencia artificial, en particular en contextos como la aplicación de la ley, puede afectar a la libertad y la seguridad de las personas y tener consecuencias de vida o muerte para quienes pertenecen a grupos raciales y étnicos marginados. El artículo 6 del Pacto destaca el derecho inherente a la vida y obliga a los Estados a proporcionar protección jurídica al respecto. El artículo 7 establece que ninguna persona será sometida a tortura u otros tratos o penas crueles, inhumanos o degradantes. El artículo 9 dispone que todo individuo tiene derecho a la libertad y a la seguridad personales y que nadie podrá ser sometido a detención o prisión arbitrarias. El artículo 14 deja claro que todas las personas son iguales ante los tribunales y cortes de justicia. El artículo 26 establece la protección de los grupos minoritarios frente a la discriminación. El párrafo 1 del artículo 2 del Pacto establece la obligación de garantizar la aplicación no discriminatoria de todas las disposiciones del Pacto. También existen disposiciones del marco del derecho internacional de los derechos humanos relativas al uso de inteligencia artificial en el control de la inmigración y las fronteras y en el contexto de las redes sociales. Estas cuestiones se analizan en informes anteriores elaborados en el marco del mandato⁸⁹.

64. El derecho internacional de los derechos humanos establece que todas las personas que puedan ser objeto de discriminación racial tienen derecho a acceder a vías de recurso

⁸⁹ A/75/590 y A/78/538.

legal, lo que se aplica en los casos en que la discriminación se produce como consecuencia de la inteligencia artificial. El artículo 6 de la Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial establece el derecho de acceso a protección y recursos efectivos ante los tribunales nacionales competentes y otras instituciones del Estado. Además, la Asamblea General ha reconocido los cinco elementos principales del derecho de las víctimas de violaciones graves de los derechos humanos a interponer recursos y obtener reparaciones: la restitución, la indemnización, la rehabilitación, la satisfacción y las garantías de no repetición⁹⁰.

65. Las empresas desempeñan un papel importante en el diseño y la aplicación de la inteligencia artificial. Son los principales agentes de su desarrollo y a menudo son contratadas por los gobiernos para implantarla en el sector público. Los Principios Rectores sobre las Empresas y los Derechos Humanos describen las obligaciones pertinentes de los gobiernos y las responsabilidades en materia de derechos humanos tanto de los gobiernos como de las empresas. Los Principios Rectores establecen que los Estados deben ofrecer protección contra los abusos de los derechos humanos cometidos por terceros, incluidas las empresas comerciales, dentro de su territorio o jurisdicción. Los Estados deben proporcionar dicha protección garantizando políticas, legislación, normativas y adjudicaciones eficaces, entre otras acciones. Los Principios Rectores establecen la responsabilidad de las empresas de prevenir, mitigar y reparar las violaciones de los derechos humanos que puedan causar o a las que puedan contribuir y de aplicar la diligencia debida en materia de derechos humanos en relación con las actividades empresariales pertinentes⁹¹. Además, los Principios Rectores establecen las obligaciones de los gobiernos y las responsabilidades de las empresas para garantizar el acceso a vías de recurso en caso de violaciones de los derechos humanos relacionadas con las empresas, complementando el derecho a la reparación previsto en otras normas, como se ha señalado anteriormente. En el marco del proyecto B-Tech del ACNUDH se han elaborado orientaciones y recursos para aplicar los Principios Rectores en el ámbito de la tecnología, incluida una labor específica sobre la inteligencia artificial⁹².

IV. Conclusiones y recomendaciones

66. **La anterior titular del mandato hizo un claro llamamiento a los Estados y otras partes interesadas, incluidas las entidades empresariales, para que se alejaran de adoptar un enfoque con “daltonismo racial” respecto a la gobernanza y la regulación de las tecnologías emergentes, incluida la inteligencia artificial. Instó a los Estados a regular estas tecnologías desde un enfoque que reconociera el racismo estructural y se basara en normas clave de derechos humanos. Sin embargo, la gestión y la regulación de la inteligencia artificial siguen siendo en gran medida insuficientes, no prestan la debida atención a los prejuicios raciales y no reflejan las normas del derecho internacional de los derechos humanos. Esta realidad persiste a pesar de la claridad y la oportunidad de los llamamientos anteriores relativos a un enfoque con “daltonismo racial” y del aumento de la concienciación sobre el racismo sistémico en los cuatro años transcurridos desde entonces. La suposición de que la tecnología es objetiva y neutra sigue estando omnipresente e impulsa una carrera por integrar la inteligencia artificial en la sociedad a pesar de sus repercusiones racialmente discriminatorias y de que no se analice en suficiente medida si es necesaria. Si bien la inteligencia artificial tiene un potencial positivo, incluso para la igualdad y la inclusión, no es la panacea para todos**

⁹⁰ Principios y Directrices Básicos sobre el Derecho de las Víctimas de Violaciones Manifiestas de las Normas Internacionales de Derechos Humanos y de Violaciones Graves del Derecho Internacional Humanitario a Interponer Recursos y Obtener Reparaciones, párrs. 15 a 23.

⁹¹ Véase también Oficina de las Naciones Unidas sobre la Prevención del Genocidio y la Responsabilidad de Proteger y Proyecto sobre Derechos Humanos, Macrodatos y Tecnología del Consejo de Investigación Económica y Social y la Universidad de Essex, “Countering and addressing online hate speech: a guide for policy makers and practitioners”, documento de política general, julio de 2023; y A/74/486, párrs. 44 y 45.

⁹² Véase ACNUDH, “B-Tech Project: OHCHR and business and human Rights”, disponible en <https://www.ohchr.org/es/business/b-tech-project>.

los problemas sociales y debe gestionarse eficazmente para lograr un equilibrio entre sus beneficios y riesgos.

67. La regulación eficaz y exhaustiva de la inteligencia artificial es fundamental para lograr este cuidadoso equilibrio. Si bien conseguir una regulación eficaz de la inteligencia artificial es vital, hay medidas adicionales que los Estados y otras instancias pueden tomar para abordar eficazmente los impactos racialmente discriminatorios de estas tecnologías. También es muy importante ofrecer a la sociedad una formación acerca de las tecnologías emergentes que esté centrada en los derechos humanos y fomentar la alfabetización en inteligencia artificial. Cuando las personas y los grupos comprenden la inteligencia artificial y son conscientes de sus derechos humanos en el espacio digital, cuentan con las facultades para utilizar ese conocimiento de forma responsable y convertirse en usuarios perspicaces y responsables que pueden mejorar la rendición de cuentas de los sistemas de inteligencia artificial.

68. Los Estados deberían:

a) Abordar el reto de regular la inteligencia artificial con un mayor sentimiento de urgencia, teniendo en cuenta la velocidad a la que se están desarrollando estas tecnologías y la multitud de formas en que ya están perpetuando la discriminación racial en todos los ámbitos sociales;

b) Desarrollar marcos reguladores de la inteligencia artificial que se basen en una comprensión global del racismo sistémico y se fundamenten en el derecho internacional de los derechos humanos, incluida la prohibición de la discriminación racial. Estos marcos no deberían basarse en enfoques aislados y deberían tener en cuenta diferentes instrumentos jurídicos, incluida la legislación específica sobre inteligencia artificial, las leyes sobre privacidad, las disposiciones sobre libertad de información, la legislación contra la discriminación y las normativas sectoriales, a fin de lograr una regulación exhaustiva y eficaz que prevenga y aborde el impacto racialmente discriminatorio de la inteligencia artificial;

c) Considerar el papel que pueden desempeñar las normas voluntarias en los marcos reguladores de la inteligencia artificial. Las normas voluntarias pueden proporcionar directrices concretas sobre la aplicación práctica de las normas jurídicas. Sin embargo, la regulación de la inteligencia artificial no debería basarse únicamente en normas voluntarias, dada la importancia de las implicaciones de estas tecnologías para los derechos humanos, entre otros en relación con la discriminación racial;

d) Consagrar la obligación jurídicamente vinculante, dentro de los marcos normativos, de llevar a cabo evaluaciones exhaustivas de la diligencia debida en materia de derechos humanos, incluidos criterios explícitos para evaluar los prejuicios raciales y étnicos, en el desarrollo y despliegue de todas las tecnologías de inteligencia artificial. La diligencia debida en materia de derechos humanos debería incluir protocolos de comprobación de datos y umbrales que protejan contra el sesgo algorítmico, incluido el sesgo racial y étnico, que deberían estar listos antes del despliegue de nuevas tecnologías, sobre todo en entornos públicos, como los educativos, los policiales y los sanitarios;

e) Considerar la posibilidad de prohibir el uso de sistemas de inteligencia artificial que hayan demostrado tener riesgos inaceptables para los derechos humanos, incluidos los que violan la prohibición de la discriminación racial;

f) Velar por que existan disposiciones en los marcos normativos que garanticen la plena transparencia sobre los procesos automatizados de toma de decisiones, incluido el derecho de acceso a la información, en los casos en que se considere admisible el uso de inteligencia artificial, sobre la base de una diligencia debida exhaustiva en materia de derechos humanos;

g) Poner en marcha procesos de apelación claros y accesibles, que estén destinados a evaluar y abordar los impactos racialmente discriminatorios de la inteligencia artificial e incluyan una revisión humana. Debería garantizarse un acceso equitativo a estos procedimientos de recurso;

h) Establecer mecanismos que permitan a las personas y grupos afectados acceder a recursos que garanticen la restitución, indemnización, rehabilitación, satisfacción y garantías de no repetición en los casos en que las tecnologías de inteligencia artificial hayan dado lugar a violaciones de los derechos humanos, incluida la discriminación racial;

i) Evitar cualquier excepción dentro de las normas reguladoras que pueda dar lugar a violaciones de la prohibición de la discriminación racial según el derecho internacional de los derechos humanos;

j) Garantizar que las partes interesadas de todos los grupos raciales y étnicos marginados, así como los profesionales de los sectores pertinentes, sean consultados de forma significativa y eficaz en el desarrollo y la aplicación de la normativa sobre inteligencia artificial, así como en el desarrollo y el uso de las tecnologías de inteligencia artificial;

k) Invertir en la recopilación de datos desglosados en todos los sectores pertinentes para obtener la información que permita que la regulación de la inteligencia artificial se base en la comprensión del racismo sistémico, que se aborden los problemas relacionados con los datos de los sistemas de inteligencia artificial y que se supervise y evalúe mejor el impacto de la tecnología de inteligencia artificial en las personas de grupos raciales y étnicos marginados;

l) Adoptar un enfoque sobre la cuestión de los datos que se fundamente en las normas de derechos humanos y posibilite el desglose, la autoidentificación, la transparencia, la privacidad, la participación y la rendición de cuentas en la recopilación y el almacenamiento de los datos⁹³;

m) Establecer mecanismos sólidos para la supervisión y el seguimiento continuo de las herramientas de inteligencia artificial, incluidas auditorías periódicas de su impacto, para garantizar el cumplimiento de la normativa y abordar cualquier preocupación planteada por las personas o comunidades afectadas, así como los posibles sesgos generados por los modelos de inteligencia artificial a lo largo del tiempo;

n) Participar en la cooperación internacional, el desarrollo de capacidades y la investigación para garantizar que los beneficios de la inteligencia artificial se distribuyan de forma más equitativa entre los países, con el fin de evitar la profundización de las desigualdades que existen como legado del colonialismo y la esclavitud;

o) Ofrecer a la sociedad una formación centrada en los derechos humanos acerca del uso aceptable y responsable de la tecnología de inteligencia artificial para aumentar la alfabetización en inteligencia artificial, con inclusión de componentes que conciencien específicamente sobre los impactos racialmente discriminatorios de la inteligencia artificial.

69. Las empresas deberían:

a) Realizar evaluaciones de diligencia debida en materia de derechos humanos en todas las fases de diseño, desarrollo e implantación de productos de inteligencia artificial;

b) Velar por el mantenimiento de consultas significativas y efectivas con personas de grupos raciales y étnicos marginados, profesionales de ámbitos sociales relevantes y expertos en racismo sistémico respecto al diseño, el desarrollo y la implantación de productos de inteligencia artificial;

c) Desarrollar protocolos con los que garantizar la plena transparencia y el intercambio de información sobre la toma de decisiones algorítmicas respecto a los productos que tienen implicaciones para los derechos humanos;

⁹³ Véanse [A/HRC/42/59](#) y [A/HRC/44/57](#).

d) Garantizar la supervisión continua de los productos de inteligencia artificial para detectar prejuicios raciales;

e) Elaborar materiales de formación sobre discriminación racial, incluidos los prejuicios implícitos y el racismo sistémico, dirigidos a todas las personas que participan en el diseño, el desarrollo y la implantación de la inteligencia artificial. La elaboración de esos materiales debería basarse tanto en las normas del derecho internacional de los derechos humanos como en las investigaciones realizadas sobre el impacto racialmente discriminatorio de las tecnologías de inteligencia artificial;

f) Contribuir a los esfuerzos por ofrecer a la sociedad una formación centrada en los derechos humanos a fin de mejorar la accesibilidad de la alfabetización en inteligencia artificial.

70. Las Naciones Unidas y sus mecanismos independientes de derechos humanos deberían:

a) Facilitar diálogos y debates efectivos sobre las tecnologías de inteligencia artificial y su regulación entre las partes interesadas;

b) Integrar un enfoque explícito sobre el impacto racialmente discriminatorio de las tecnologías de inteligencia artificial en la labor del Órgano Asesor de Alto Nivel sobre Inteligencia Artificial;

c) Garantizar que las publicaciones y orientaciones sobre tecnologías de inteligencia artificial se basen en el derecho internacional de los derechos humanos, incluida la prohibición de la discriminación racial, y reconozcan explícitamente el sesgo racial en el diseño y despliegue de la tecnología de inteligencia artificial como un grave problema mundial;

d) Desempeñar un papel en el seguimiento de las implicaciones para los derechos humanos de las tecnologías de inteligencia artificial, incluidas las relacionadas con la discriminación racial;

e) Apoyar la cooperación internacional, la creación de capacidad y la investigación para tratar de lograr una distribución más equitativa de los beneficios de la inteligencia artificial entre los países.
