

**Aportación técnica en el marco de la
audiencia temática regional de la
Comisión Interamericana de Derechos
Humanos**

**“Inteligencia Artificial y Derechos
Humanos”**

Sometida por Electronic Frontier Foundation y
ARTICLE 19 Oficina para México y
Centroamérica

Sobre Artículo 19 Oficina para México y Centroamérica

ARTICLE 19 Oficina para México y Centroamérica (ARTICLE 19) es una organización no gubernamental independiente que promueve que todas las personas puedan ejercer su derecho a informarse y expresarse en un entorno de libertad, seguridad e igualdad tanto en el ámbito *offline* como *online*.

La organización inició hace más de cuatro años un trabajo en la agenda de inteligencia artificial en México. La intención fue entender cómo se originaron los esfuerzos incipientes en su desarrollo, implementación y regulación, los obstáculos y retos que se presentaron y el grado de voluntad política para adoptar dicha agenda como políticamente prioritaria -dada la evolución de las tecnologías y la incorporación de éstas en distintos ámbitos de la vida pública-. Los hallazgos y conclusiones de nuestra exploración inicial revelaron un vacío y desinterés por discutir el tema de tecnologías emergentes en manos del sector público con la seriedad que el tema exige. Sin embargo, recientemente la nueva administración gubernamental mexicana ha dejado claro un interés por abordar el tema tecnológico, iniciando por reducir la brecha digital a través de infraestructura, y paralelamente apostando por la digitalización al sector público para llevar a cabo distintos procesos, por ejemplo, los trámites administrativos y la impartición de justicia.

Sin embargo, desde ARTICLE 19 hemos observado actualmente que existe una falta de comprensión de lo que es la inteligencia artificial y el impacto que puede causar en los derechos humanos. Esto sucede cuando existen iniciativas legislativas que inhiben la interdependencia de los derechos humanos. Además, ARTICLE 19 ha analizado la carencia de profundidad y entendimiento de la transparencia algorítmica, resaltando que es necesario abordar su estructura, retos y la prioridad de vincularla con principios y lineamientos relacionados con el acceso a la información, la transparencia y la rendición de cuentas. Lo anterior no solo dirigido a las empresas privadas que desarrollan estas herramientas, sino también a los gobiernos que están utilizando estas tecnologías dentro de su estrategia nacional de digitalización.

En este sentido, ARTICLE 19 ha sido invitada como organización experta para participar y facilitar el debate con otras partes interesadas, contribuyendo a estas conversaciones con una perspectiva de derechos humanos y centrándose en cuestiones de libertad de expresión y acceso a la información. ARTICLE 19 trabaja para vincular la promoción de políticas públicas, el acompañamiento a procesos locales y el litigio estratégico para promover los más altos estándares internacionales de derechos humanos en el desarrollo, uso, adquisición e implementación de herramientas de tecnologías emergentes, de tal manera que se contribuya al fortalecimiento de la democracia.

Sobre la Electronic Frontier Foundation

La Electronic Frontier Foundation (EFF) es una organización no gubernamental sin ánimo de lucro con sede en Estados Unidos y que opera a escala internacional para la protección de los derechos humanos y las libertades fundamentales en nuestra era digital. Fundada en 1990, la EFF defiende la privacidad, la libertad de expresión y la innovación mediante litigios estratégicos, análisis de políticas públicas, activismo de base y desarrollo tecnológico. La misión de la EFF es garantizar que la tecnología apoye las libertades, la justicia y la innovación para todas las personas.

Como parte de su labor, la EFF viene siguiendo iniciativas de regulación de la inteligencia artificial en América Latina y el despliegue de estas tecnologías para utilizaciones sensibles, en particular por instituciones estatales para actividades y determinaciones que afectan derechos humanos. La falta de transparencia significativa y participación ciudadana en la forma en que en general se desarrollan, implementan y utilizan estos sistemas por las instituciones estatales son factores clave que plantean graves desafíos a la gobernanza democrática, a la legitimidad de la toma de decisiones y la formulación de políticas gubernamentales. En base a los estándares interamericanos de derechos humanos, hemos desarrollado implicaciones que los Estados deben tener en cuenta desde el momento en que analizan si es adecuado adoptar un sistema algorítmico como parte de determinada acción estatal hasta el momento en que evalúan los resultados de su implementación.

En relación con la audiencia temática regional titulada *Inteligencia artificial y derechos humanos* que se realizó el viernes 7 de marzo de 2025, ARTICLE 19 y EFF someten los presentes aportes técnicos con el propósito de analizar los retos que la inteligencia artificial (IA) plantea para el ejercicio y garantía de los derechos humanos, y ofrecer algunas recomendaciones para prevenirlos y mitigarlos.

A. La inteligencia artificial y su relación con el ejercicio y la garantía de protección de los derechos humanos

1. Moderación automatizada de contenidos en plataformas de redes sociales

La IA se ha posicionado como un catalizador de la transformación¹ en todas las disciplinas e industrias, reconfigurando procesos tradicionales. Es por ello por lo que la IA se ha integrado en diversas disciplinas y prácticas no solo jurídicas sino temas como la salud, la educación, la geografía y muchas otras. Por ejemplo, en el ámbito académico, la IA puede potencializar la investigación y la docencia, facilitando el análisis de grandes volúmenes de información y optimizando la elaboración de texto jurídicos.

Internet ya no es solo un medio para difundir y publicar información, es también una herramienta necesaria para el ejercicio de los derechos humanos. Es así como la tecnología se convierte más en una herramienta necesaria como parte de la vida laboral, familiar y social, de la educación, de la expresión y de la construcción de nuestra identidad en la esfera colectiva como en la individual.

Un ejemplo de ello es como las redes sociales se han convertido en espacios de discusión, de intercambio de conocimientos, de reflexiones, de pugnas e incluso de contiendas narrativas en foros donde muchas personas discuten sobre diversas temáticas.

Como cualquier otro espacio social, Internet ha también cambiado las dinámicas de interacción virtual donde también se reproducen estereotipos, roles y prácticas sociales del espacio offline. Muestra de ello como la violencia, la desigualdad y la injusticia se refleja en los espacios digitales, así como las brechas económicas, educativas y sanitarias de la desigualdad por razones de género, raciales, étnicas, de clase social que crecen y se profundizan en nuestras comunidades.

Debido a algunas prácticas dadas tanto en las redes sociales y plataformas digitales se presenciado prácticas de violencia desmedida además de profundizar la desigualdad, las plataformas digitales han creado mecanismos de control o de moderación de contenido al clasificar bajo criterio propio aquellos discursos, contenido audio-visual y expresiones como nocivo y/o problemático cuando infringen las normas, reglamentos o términos de servicio de su plataforma.

¹ <https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf>

2. ¿La IA como una herramienta para potencializar la democracia? Riesgos y oportunidades de la IA

Las tecnologías de la información (TI) en el contexto gubernamental son fundamentales para impulsar la eficiencia operativa y facilitar la toma de decisiones basada en datos. Comprenden una gama de soluciones, incluyendo sistemas computacionales, software y redes, que se utilizan para procesar y distribuir información de manera efectiva. Estas tecnologías permiten a los gobiernos mejorar la prestación de servicios públicos al automatizar tareas, grandes volúmenes de información y comunicar datos.

B. Identificación y prevención de impactos diferenciados en grupos vulnerables

El uso de IA en grupos en situación de vulnerabilidad y marginación

El informe [Libertad artificial: moderación de contenidos y su impacto en derechos humanos](#), elaborado por ARTICLE 19², aborda cómo la moderación automatizada de contenidos en plataformas digitales puede afectar el derecho a la libertad de expresión y otros derechos fundamentales. Se basa en un análisis documental, encuestas, entrevistas y talleres con comunidades vulnerables en México.

El estudio destaca que la moderación de contenido realizado por plataformas digitales, en particular cuando se realiza de manera automatizada, puede dar lugar a censura, discriminación y autocensura, impactando de manera diferenciada a distintos sectores de la sociedad que temen [represalias](#) de actores gubernamentales y/o no estatales³. Con relación a lo anterior, las plataformas digitales no siempre logran controlar eficazmente los denominado discursos de odio, lo que afecta especialmente a mujeres, periodistas, defensores de derechos humanos, personas de LGTBQ+ y migrantes, quienes son más propensos a sufrir acoso y violencia digital. De esta manera las personas con discapacidad en la mayoría de las ocasiones son estigmatizadas y sus contenidos suelen ser malinterpretados o eliminados por los algoritmos.

Aunado a lo anterior, el informe muestra evidencia sobre cómo la moderación automatizada de contenido en plataformas digitales puede reforzar desigualdades estructurales y restringir derechos fundamentales como la libertad de expresión y acceso a la información. En

² ARTICLE 19 Oficina México y Centroamérica. *Libertad Artificial: moderación de contenidos y su impacto en derechos humanos*. 19 de octubre de 2022. <https://articulo19.org/libertad-artificial-discurso-redes-y-pluralidad/>

³ ARTICLE 19 Oficina México y Centroamérica. *Derechos Pendientes*. 9 de julio de 2024. <https://articulo19.org/derechospendientes/>

conclusión, dicho documento hace un llamado urgente a garantizar la transparencia, mecanismos de apelación y una regulación basada en derechos humanos para asegurar que las plataformas digitales no perpetúan la discriminación y la censura en el espacio digital.

Algunos de los impactos que se destacan son:

1) Libertad de expresión y acceso a la información

- La eliminación de contenido legítimo limita la diversidad de voces en el espacio digital.
- Se restringe el debate público en temas clave como derechos humanos, política y activismo social.
- Se genera un efecto de autocensura: las personas dejan de expresar ciertas opiniones por temor a ser bloqueadas.

2) Derechos de grupos vulnerables

- Comunidades indígenas y afromexicanas: Existen barreras idiomáticas y culturales que impiden la correcta interpretación de su contenido por parte de los algoritmos.
- Personas LGBTQ+: Han enfrentado eliminaciones injustificadas de contenido relacionado con su identidad y derechos.
- Mujeres: Se ha detectado la invisibilización de denuncias sobre violencia de género, mientras que los ataques digitales en su contra persisten sin una respuesta adecuada

3) Privacidad y seguridad digital

- Las plataformas recolectan grandes volúmenes de datos sin garantizar transparencia en su uso.
- No existen garantías claras sobre cómo se almacenan o utilizan los datos generados por los usuarios en procesos de moderación de contenido

Uso de tecnologías emergentes y ausencia de transparencia algorítmica en sector público

ARTICLE 19 publicó el informe titulado [“Hacia una transparencia algorítmica en el sector público de América Latina”](#) la cual aborda los desafíos de la transparencia en la adquisición, desarrollo e implementación de tecnologías emergentes en la administración pública de Argentina, Brasil, Colombia, Chile y México. El estudio identifica la opacidad del uso de algoritmos y sistemas de inteligencia artificial (IA) en el sector público donde se destacan tres problemas fundamentales:

- A. **Falta de regulación sobre la transparencia algorítmica:** No existen normativas específicas en la región que obliguen a los gobiernos a explicar cómo funcionan los sistemas de IA utilizados en la toma de decisiones administrativas;
- B. **Dificultad en la supervisión y rendición de cuentas:** La administración pública carece de mecanismos claros para monitorear el impacto de la IA y garantizar que sus decisiones sean auditables;
- C. **Impacto en derechos humanos y participación ciudadana:** La falta de transparencia en la implementación de tecnologías emergentes impide que la ciudadanía pueda cuestionar o apelar decisiones automatizadas.

Uno de los principales hallazgos se relaciona con la falta de acceso a información sobre IA en el sector público, pues es relevante destacar que en ninguno de los cinco países estudiados existe un registro centralizado de los sistemas de IA utilizados en la administración pública. Aunado a ello, los contratos y licitaciones vinculadas con tecnologías emergentes suelen estar protegidos por acuerdos de confidencialidad, lo cual impide su conocimiento. Para este tipo de situaciones no se encontraron estándares claros para garantizar que los algoritmos utilizados sean explicables y accesibles para la ciudadanía.

Compra, adquisición y uso de tecnologías emergentes por el sector público (cadena de suministro)

La adquisición y el despliegue de sistemas de IA por parte de los Estados presentan riesgos significativos para los derechos humanos en diversos ámbitos, con profundas implicaciones para la privacidad, la igualdad, la responsabilidad democrática y la participación pública en los sistemas de innovación digital.

(i) Aplicación de la ley: El uso de la IA en la aplicación de la ley, como la policía predictiva y la tecnología de reconocimiento facial, plantea importantes problemas de derechos humanos, incluidos los resultados sesgados que perjudican desproporcionadamente a las comunidades marginadas, conducen a encarcelamientos injustos y erosionan la privacidad de las personas a través de una amplia recopilación de datos.⁴ Aunque tecnologías como el reconocimiento facial pueden presentar una tasa de error baja en términos absolutos, su despliegue sigue dando lugar a que se rastree erróneamente a un gran número de personas, lo que amplifica la vigilancia de los movimientos y las vidas de las personas. Estas prácticas amenazan la autonomía individual, ahogan la libertad de expresión y amenazan con crear un Estado de

⁴ <https://link.springer.com/article/10.1007/s13347-021-00478-z>

vigilancia.⁵ La opacidad de estos sistemas de IA dificulta aún más la rendición de cuentas de las fuerzas del orden.

(ii) Asistencia social: Los sistemas de asistencia social basados en la IA corren el riesgo de exclusiones injustas, vigilancia y falta de rendición de cuentas.⁶ La denegación arbitraria e injusta de prestaciones sociales⁷ priva a las personas del derecho a la vida, a la igualdad, a un nivel de vida adecuado y a la subsistencia.⁸ Cuando los servicios públicos, como la asignación de recursos, no son transparentes, se limitan las oportunidades de participación pública en el desarrollo y la evaluación de políticas, obstaculizándose también la capacidad de los sujetos para solicitar reparación en caso de parcialidad u otros perjuicios. Esto ocurre especialmente cuando intervienen agentes privados en el desempeño de funciones de servicio público.

(iii) Educación: La IA en la educación plantea riesgos como la vigilancia omnipresente, las recompensas o castigos basados en el comportamiento y la asfixia del pensamiento crítico.⁹ Puede exacerbar la violencia, el acoso y el ciberacoso.¹⁰ Un Relator Especial de la ONU destaca los riesgos de utilizar la IA para las admisiones, la asignación de alumnos, las evaluaciones de resultados y la supervisión del comportamiento durante los exámenes.¹¹ La IA también podría amplificar las desigualdades educativas, aumentando las diferencias entre estudiantes privilegiados y desfavorecidos, ya que las herramientas de IA a menudo favorecen las culturas e idiomas dominantes.¹² También suscitan preocupación la reducción de la interacción humana, el declive del papel de los profesores, los enfoques tecnocráticos y la explotación de los datos de los alumnos con fines no consentidos o socialmente hostiles.¹³

(iv) Servicios de protección de la infancia: Los Estados automatizan cada vez más los procesos de predicción del riesgo de violación de los derechos de los niños y adolescentes. El creciente uso de la IA para predecir los riesgos para los derechos de los niños permite a los actores del sector privado acceder y explotar datos altamente sensibles sin protecciones adecuadas de la privacidad o salvaguardias contra el abuso y la explotación.¹⁴

5 <https://articulo19.org/continua-impune-el-uso-de-pegasus-a-un-ano-de-las-nuevas-denuncias/>

6 <https://www.amnesty.org/en/latest/news/2024/11/denmark-ai-powered-welfare-system-fuels-mass-surveillance-and-risks-discriminating-against-marginalized-groups-report/>

7 <https://www.amnesty.org/en/latest/research/2024/04/entity-resolution-in-indias-welfare-digitalization/>

8 <https://ohrh.law.ox.ac.uk/the-human-rights-implications-of-chinas-social-credit-system/>

9 <https://www.mdpi.com/2071-1050/15/16/12451>

10 <https://slejournal.springeropen.com/articles/10.1186/s40561-024-00316-7>

11 <https://documents.un.org/doc/undoc/gen/n24/298/43/pdf/n2429843.pdf>

12 https://www.researchgate.net/publication/351119867_Equity_and_Artificial_Intelligence_in_Education_Will_AIEd_Amplify_or_Alleviate_Inequities_in_Education

13 <https://documents.un.org/doc/undoc/gen/n24/298/43/pdf/n2429843.pdf>

14 <https://www.mdpi.com/2071-1050/15/16/12451>

(v) Poder judicial: Dado que la IA tiende a reforzar los prejuicios estructurales, su uso en el poder judicial, especialmente para la evaluación de riesgos en las sentencias, puede producir resultados discriminatorios,¹⁵ socavando el derecho a un juicio y un trato justos, así como comprometiendo la discreción judicial.¹⁶ Una reciente sentencia dictada por un juez en Colombia para proteger el derecho a la salud de un menor con autismo, en la que el razonamiento jurídico fue redactado íntegramente por ChatGPT suscita preocupación en relación con la independencia judicial y el deber de dictar sentencias motivadas.¹⁷

(vi) Migración y gestión de fronteras: Las tecnologías de IA en los sistemas de migración amenazan cada vez más los derechos humanos de los migrantes y su acceso al asilo.¹⁸ Un informe de la Electronic Frontier Foundation de 2024 destaca cómo las tecnologías de vigilancia en capas, con frecuencia automatizadas por IA, controlan las comunidades fronterizas.¹⁹ La Plataforma Centinela en Chihuahua demuestra cómo los sistemas opacos, que incluyen la identificación biométrica y la predicción del comportamiento, se utilizan con el pretexto de la seguridad pública y el control de la inmigración, lo que plantea preocupaciones sobre la transparencia, la rendición de cuentas y los impactos en los derechos humanos.

(vii) Guerra y conflicto: La IA en aplicaciones militares plantea graves riesgos para la vida civil y la rendición de cuentas. Las armas autónomas y los sistemas de vigilancia basados en IA reducen la supervisión humana y difuminan la rendición de cuentas.²⁰ En 2024, expertos de la ONU condenaron el uso de IA²¹ y pidieron medidas reparadoras, destacando las implicaciones éticas y humanitarias de tales tecnologías.²²

Los ejemplos anteriores demuestran cómo las funciones críticas del Estado, como el bienestar, la educación, la justicia y el mantenimiento de los bienes comunes públicos, están amenazadas de erosión por el lento avance de los intereses privados, que no rinden cuentas, en el ámbito público.²³ La fuga del valor de los datos de la esfera pública a ecosistemas

15 <https://researchoutreach.org/articles/justice-served-discrimination-in-algorithmic-risk-assessment/>

16 https://www.unodc.org/dohadeclaration/en/news/2021/30/judge-dread_ai-and-judicial-integrity.html

17 <https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling>

18 <https://www.amnesty.org/en/latest/news/2024/05/global-new-technology-and-ai-used-at-borders-increases-inequalities-and-undermines-human-rights-of-migrants/>

19 <https://www.eff.org/files/2024/05/06/borderzine-2024-5-6-es.pdf>

20 <https://hms.harvard.edu/news/risks-artificial-intelligence-weapons-design>

21 Domicidio se refiere a la violación sistemática o generalizada del derecho a una vivienda adecuada; Véase, <https://documents.un.org/doc/undoc/gen/n22/431/55/pdf/n2243155.pdf>

22 <https://ainowinstitute.org/publication/the-algorithmically-accelerated-killing-machine>

23 <https://itforchange.net/sites/default/files/2647/Recovering%20the%20Public%20in%20India%E2%80%99s%20Digital%20Public%20Infrastructure%20Strategy.pdf>

cautivos del sector privado apunta a la necesidad de democratizar los datos y los dividendos de la IA para que sirvan al bien colectivo.

Ahora bien, en el caso de México ante la desaparición del Instituto de Acceso a la Información y Transparencia (INAI) muestra una clara debilidad institucional ante un riesgo inminente para acceder a la información en donde el 90% de las adquisiciones tecnológicas se realizan por adjudicación directa, limitando la competencia y aumentando riesgos de corrupción e incluso abriendo la posibilidad de cometerse violaciones a derechos humanos.

Caso Fiscalía de Nayarit, México. Diagnóstico de uso e implementación de tecnologías de IA en la impartición de justicia penal.

El documento titulado [“Reflexiones sobre la implementación de sistemas de IA en la procuración de justicia. Estudio de caso: Nayarit”](#) analiza los desafíos, oportunidades y riesgos de integrar la inteligencia artificial en el sistema de justicia penal, particularmente en México, tomando como ejemplo el estado de Nayarit.²⁴ Dicho informe propone un marco para implementar la IA en la procuración de justicia de manera ética, transparente y respetuosa de los derechos humanos donde se identifican áreas críticas de mejora.

El documento menciona desde la atención a la infraestructura tecnológica deficientes (hasta 26.32%) hasta la falta de capacitación del personal (incluye un 87.72%) la cual puede aparejarse riesgos como los sesgos en los algoritmos, problemas de transparencia y la protección de datos personales donde el 98.25% del personal ignora los derechos ARCO (acceso, rectificación, cancelación y oposición) y la gestión inadecuada de archivos y ciberseguridad donde el 77.19% del personal desconoce sistemas de ciberseguridad, además de la existencia de prácticas inseguras como compartir contraseñas y usar dispositivos externos sin control.

Los hallazgos incluyen que la mayoría de las personas no está capacitada en IA, existen equipos obsoletos y no se cumplen con las normativas de protección de datos. En ella también se menciona que el 82.45% de las personas no usa herramientas de colaboración en línea, eso indica una brecha digital importante del 94.74%. Aunado a lo anterior, la falta de documentos clave como avisos de privacidad y análisis de riesgos muestra un incumplimiento de leyes, lo que podría llevar a sanciones o incluso perpetuarse violaciones derechos humanos.

²⁴ <https://articulo19.org/caso-nayaritreflexiones-sobre-la-implementacion-de-sistemas-de-ia-en-la-procuracion-de-justicia/>

El impacto sobre los derechos humanos que se evaluó con la información recabada fue el siguiente:

1. Privacidad y protección de datos:

- La ausencia de medidas robustas expone datos personales de víctimas, testigos y acusados a filtraciones o uso indebido, violando el derecho a la privacidad (Art. 16 constitucional).

2. No discriminación y igualdad:

- Sesgos en algoritmos podrían replicar patrones discriminatorios (ej: perfiles raciales o socioeconómicos), afectando a grupos vulnerables y violando el principio de igualdad (Art. 1º constitucional).

3. Debido proceso:

- Decisiones automatizadas sin supervisión humana podrían generar errores en imputaciones o medidas cautelares, comprometiendo el derecho a un juicio justo (Art. 20 constitucional).

4. Acceso a la justicia:

- La falta de transparencia en el uso de IA erosiona la confianza ciudadana, dificultando el acceso efectivo a la justicia.

C. La implementación de mecanismos de protección efectivos

Los mecanismos estatales que abordan las repercusiones negativas de la IA sobre los derechos humanos se dividen generalmente en tres categorías:

(i) Evaluaciones de impacto sobre los derechos humanos (EIDH):

Las EIDH por parte de los Estados y las empresas pueden servir como herramienta de empoderamiento, dotando a las personas afectadas de información para cuestionar el diseño, despliegue y funcionamiento de los sistemas de IA. Por ejemplo, el artículo 72 de la Ley de AI de la UE obliga a realizar evaluaciones de impacto sobre los derechos fundamentales para los sistemas de AI de alto riesgo, lo que permite la transparencia y la rendición de cuentas.

(ii) Notificación y audiencia justa antes de decisiones adversas:

Las personas deben ser informadas cuando se utilice un sistema automatizado en decisiones que les afecten. El Blueprint for the AI Bill of Rights de EE.UU. hace hincapié en la notificación oportuna y comprensible, las opciones de exclusión voluntaria y las explicaciones sobre

cambios funcionales clave.²⁵ Las personas también deben tener acceso a mecanismos humanos de supervisión y apelación antes de que se tomen decisiones adversas. Estos mecanismos deben ser accesibles, equitativos, eficaces y oportunos, y no deben imponer una carga excesiva a la persona afectada. Otros ejemplos son la Carta de Algoritmos de Nueva Zelanda, que obliga a recurrir las decisiones automatizadas,²⁶ y las leyes estadounidenses de curación de papeletas de 24 Estados, que ofrecen sistemas alternativos para los votantes marcados por algoritmos de cotejo de firmas.²⁷

(iii) Repositorios de algoritmos públicos: Dentro del sector público, la creación de repositorios de algoritmos públicos puede mejorar la explicabilidad y la rendición de cuentas al abrir la caja negra del desarrollo y la implantación de la IA. Incluso cuando la transparencia total de los algoritmos puede no ser deseable, los repositorios pueden contener casos de uso en los que se utilizan los algoritmos, junto con explicaciones de su papel en el proceso de toma de decisiones, garantizando la transparencia y permitiendo la reparación de los impactos adversos.²⁸ Estos repositorios también pueden formar parte de una estrategia de transparencia algorítmica más amplia, que se extienda al sector privado e integre la EIDH, la reparación de agravios y los requisitos de información.

(iv) Recursos judiciales: Los tribunales pueden desempeñar un papel fundamental a la hora de abordar las injusticias relacionadas con la IA. Para mejorar el acceso a la justicia, los organismos de igualdad y las organizaciones de interés público deben estar facultados para presentar denuncias, incluso sin denunciadores identificables, ya que las violaciones de derechos impulsadas por la IA suelen pasar desapercibidas para las víctimas.

La revisión de las normas sobre pruebas también es crucial para aliviar la carga de la prueba de los demandantes, dada la complejidad y opacidad de los sistemas de IA. La Directiva de la UE sobre responsabilidad por IA propone una presunción de causalidad entre el incumplimiento y el daño.²⁹

Ejemplos de casos judiciales:

- **El caso SyRI holandés (2020):** El Tribunal de Distrito de La Haya dictaminó que el sistema de algoritmos System Risk Indication (SyRI), un instrumento legal que el

²⁵ <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>

²⁶ http://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf

²⁷ <https://www.lawfareblog.com/mail-voting-litigation-2020-part-iv-verifying-mail-ballots>

²⁸ <https://wp.oecd.ai/app/uploads/2024/12/16-Algorithmic-Transparency-in-the-Public-Sector-Recommendations-for-Governments-to-Enhance-the-Transparency-of-Public-Algorithms.pdf>

²⁹ [https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI\(2023\)739342_ES](https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI(2023)739342_ES).

gobierno holandés utiliza para detectar el fraude en ámbitos como las prestaciones, los subsidios y los impuestos, viola el derecho a la intimidad en virtud del Convenio Europeo de Derechos Humanos.³⁰

- **Mobley contra Workday (2023):** Un tribunal estadounidense admitió a trámite una demanda contra las herramientas de contratación basadas en IA de Workday, que supuestamente discriminaban por motivos de raza, edad y discapacidad. La sentencia puso de relieve que los proveedores de IA pueden ser considerados responsables en virtud de las leyes contra la discriminación cuando actúan como agentes de los empleadores.³¹

D. Cumplimiento de las obligaciones estatales en base a estándares interamericanos.

Mirando los compromisos de los Estados de la región frente a los derechos humanos, la Electronic Frontier Foundation (EFF) publicó recientemente [un informe](#) sobre el uso estatal de sistemas basados en algoritmos e inteligencia artificial (IA) para automatizar tareas y procedimientos en apoyo a funciones y decisiones que afectan derechos. El informe destaca que, aunque el despliegue de estas tecnologías por entes públicos pueda ser novedoso, la base de garantías de derechos sobre la cual se debe asentar su utilización viene de muchos años y cuenta con estándares robustos que los Estados deben garantizar.

Debe recordarse que la Convención Americana sobre Derechos Humanos establece a los Estados Parte [la obligación](#) de respetar y garantizar los derechos humanos, los cuales incluyen los deberes de prevenir, investigar y castigar las violaciones a dichos derechos. En este marco, y teniendo en cuenta secciones anteriores de esta aportación, es clave recalcar que la adopción de sistemas algorítmicos como parte de políticas o iniciativas de órganos del Estado implica riesgos de seguridad, discriminación, vigilancia ilegal y errores que pueden afectar el acceso a servicios esenciales o conllevar a consecuencias graves sobre la vida de miles de personas.

La implementación de tecnología en el sector público requiere entonces de un examen detenido de los órganos que contemplen la adquisición o uso de estas tecnologías para conocer si el uso de un sistema de IA es adecuado para el propósito pretendido, así como si presenta un beneficio concreto. Estos propósitos y beneficios deben estar asociados al fortalecimiento en la garantía y el ejercicio de determinados derechos, lo que se debe comprobar al momento de monitorear y evaluar su implementación. Es así, que la adopción

³⁰<https://www.loc.gov/item/global-legal-monitor/2020-03-13/netherlands-court-prohibits-governments-use-of-ai-software-to-detect-welfare-fraud/>

³¹<https://www.seyfarth.com/news-insights/mobley-v-workday-court-holds-ai-service-providers-could-be-directly-liable-for-employment-discrimination-under-agent-theory.html>

de estos sistemas por entes estatales demanda la realización de evaluaciones de impacto de derechos humanos antes del despliegue y durante su uso.

Aunado a ello, cualquier violación de derechos humanos derivada de la adopción de estos sistemas y tecnologías por instituciones públicas debe resultar en investigaciones que proporcionen insumos para la evaluación del sistema y del contexto de su utilización, conduciendo a modificaciones efectivas o a la interrupción en su uso. Con ello, también se desarrollan las medidas necesarias para evitar que la violación de derechos se repita.

Respetar y garantizar derechos implica organizar todo el aparato del poder público para hacer efectivos los derechos establecidos por la Convención, incluyendo adoptar medidas legislativas y ajustar prácticas, procesos y estructuras del Estado hacia este objetivo. Todas las instituciones del Estado están sujetas a los compromisos asumidos frente al derecho internacional de derechos humanos. Esto incluye a los jueces y tribunales a los cuales compete ejercer un control de convencionalidad sobre las leyes y reglamentos nacionales que basan el uso estatal de sistemas de IA o decisiones automatizadas en situaciones o mediante procedimientos ilícitos según la Convención Americana.

Principios transversales

Un elemento primordial que se desprende de las garantías interamericanas es el imperativo de la transparencia de las actividades estatales. A ellas se aplica el principio de la máxima divulgación, según el cual el derecho de acceso a la información es la regla y el secreto es la excepción. Corresponde al Estado la carga de la prueba de que el acceso a determinada información puede ser legítimamente restringido. Además, el derecho a la información tiene prioridad frente a normas contradictorias o la falta de regulación.

No se permiten restricciones amplias basadas en una calificación general de “orden público”, “seguridad nacional” o “seguridad pública”. Las limitaciones deben aplicarse estrictamente a los tipos de información y a los tipos de divulgación que, conforme a la ley, representen efectivamente un riesgo según los principios democráticos, y solamente hasta que exista dicho riesgo específico y objetivo. La exigencia de transparencia se refuerza en iniciativas que involucran altos montos del presupuesto público, lo que suele ser la regla en el despliegue de tecnologías de IA por entes estatales.

Las garantías de debido proceso en la implementación de tecnologías de IA deben ser parte de las bases esenciales de su adopción para la toma de decisiones que afectan derechos. Las personas afectadas por decisiones automatizadas deben ser informadas de los procedimientos que analizan sus derechos y obligaciones antes de que se adopte una decisión. Ellas deben recibir, con la antelación suficiente para que puedan intervenir,

información clara que explique por qué están sujetas al procedimiento, los elementos relevantes que se están considerando y una referencia sobre cómo los elementos que se están evaluando influyen en las consecuencias a las que pueden enfrentarse. La información necesaria incluye aún los medios de que ellas disponen para presentar sus alegaciones, que deben ser fácilmente accesibles. Además, las decisiones deben ser debidamente justificadas, lo que implica a los sistemas utilizados observar estándares de interpretabilidad y explicabilidad.

Se suman a estas garantías el derecho a la vida privada en su relación con la protección de datos y la autodeterminación. En una [decisión reciente](#) de la Corte Interamericana de Derechos Humanos se ha reconocido la protección de la autodeterminación informativa bajo la Convención Americana. Con ello, se garantiza una serie de derechos vinculados al poder asegurado a las personas de conocer y tener injerencia sobre cómo sus datos personales son tratados por terceros.

Es fundamental recordar aún que la Convención Americana prohíbe todo trato discriminatorio. En consecuencia, antes de la implementación del sistema algorítmico y durante todo su ciclo de vida, los Estados deben adoptar todas las medidas necesarias para garantizar que no den lugar a un trato discriminatorio, es decir, a cualquier exclusión, restricción o privilegio que no sea objetivo, necesario y proporcional al fin legítimo que pretenden alcanzar y que afecte negativamente a los derechos humanos. Cumple así a los Estados adoptar prácticas para cumplir esta obligación. Dichas prácticas incluyen la ya mencionada realización de evaluaciones de impacto a derechos humanos y otras medidas como: garantizar que las instituciones estatales responsables y los órganos de supervisión tengan acceso al funcionamiento interno del sistema y los conocimientos adecuados para evaluarlo desde la perspectiva de los derechos humanos; realizar auditorías independientes internas y externas; y garantizar que exista una supervisión humana y una revisión humana significativas de las decisiones basadas en IA por parte de funcionarios competentes siguiendo criterios estrictos y transparentes.

La evaluación periódica, las evaluaciones de impacto y las auditorías deben abordar no solo el diseño inicial, sino los resultados reales del uso de sistemas algorítmicos una vez implementados. Las evaluaciones de impacto y las auditorías en materia de derechos humanos deben adoptar una perspectiva de género e interseccional en observación de los tres pasos que la CIDH ha señalado para el diseño de políticas públicas con enfoque de derechos humanos³². Estos pasos configuran una hoja de ruta pertinente para que los Estados evalúen las políticas públicas que hacen uso de los sistemas algorítmicos. Ellos abarcan la consideración de: (i) el impacto diferencial que una determinada política tiene o puede tener para los grupos históricamente discriminados; (ii) las perspectivas y preocupaciones de estos

³² CIDH, Políticas Públicas con Enfoque de Derechos Humanos, 15 de septiembre de 2018, para. 48.

grupos en las diferentes etapas del ciclo de la política; (iii) y los beneficios reales que esta política aporta o puede aportar para reducir las desigualdades que afectan a dichos grupos.

Se debe recalcar que corresponde a las autoridades estatales, y no a las personas afectadas, justificar y demostrar que una decisión basada en sistemas automatizados o modelos de IA no tuvo un propósito o efecto discriminatorio. Para cumplir adecuadamente con esta obligación, los Estados deben garantizar que las decisiones basadas en estos sistemas tienen razones explicables y justificables. Los mecanismos de participación pueden adoptar formas múltiples y complementarias. El [marco operativo](#) elaborado por la EFF en su informe articula algunos de estos mecanismos y estructuras, incluyendo la auditoría del sistema en colaboración con científicos de datos independientes y expertos de la sociedad civil, procesos de consulta más amplios, mecanismos de retroalimentación y queja, participación ciudadana dentro de las instituciones de supervisión, bien como la coordinación con órganos públicos que abogan por el interés de la sociedad (como Defensorías del Pueblo, órganos de defensa del consumidor, entre otros).

Finalmente, la participación social significativa es otro elemento clave de una actuación estatal basada en los derechos humanos. Ello presupone el acceso a la información relevante sobre el uso de estos sistemas y el involucramiento de distintos sectores de la sociedad en los momentos de decisión y evaluación acerca de su implementación, especialmente de los grupos históricamente discriminados, personas afectadas y expertos en los temas vinculados a cada una de estas iniciativas (ej. educación, protección social, etc.).

La transparencia y acceso a la información, el debido proceso, la igualdad y no discriminación, la privacidad y protección de datos, así como la participación social significativa son todos principios transversales que deben guiar la actuación del Estado en el despliegue de sistemas de IA en el desarrollo de sus funciones—no solamente para orientar **cómo** hacerlo y evaluar su implementación pero también para basar la definición primera de **si** dicho despliegue es legítimo y conforme a los derechos humanos.

Líneas rojas y test tripartito para la restricción y limitación de derechos humanos

Es crucial remarcar que el despliegue de sistemas algorítmicos para apoyar el desempeño de funciones públicas no es siempre un enfoque adecuado para enfrentar los problemas que se busca abordar. Si bien el uso por parte de los Estados de los sistemas algorítmicos puede en teoría servir para impulsar el ejercicio de los derechos humanos, apoyar la toma de decisiones sensibles de la función estatal en estos sistemas conlleva, al menos, la posibilidad de interferir en uno o más de esos derechos y libertades. En consecuencia, el desarrollo o la adopción de estos sistemas por parte de los Estados debe respetar los fundamentos que la Convención establece para la restricción o limitación admisibles de los derechos humanos, que es el test tripartito. Esto significa que, además de una base jurídica adecuada, las restricciones o limitaciones deben perseguir un fin legítimo y ser adecuadas, necesarias y proporcionadas

para alcanzar dicho fin, garantizando que se cumpla plenamente el deber de protección del Estado.

En el caso de los derechos sociales, económicos y culturales (DESC), dicho análisis debe integrar los principios de desarrollo progresivo y no regresión. La adopción por parte de los Estados de sistemas algorítmicos en el contexto de la protección social, o bienestar, debe favorecer el adecuado disfrute y consecución de los DESC, y no debe implicar una regresión en las garantías y políticas de protección social. Cualquier restricción a los DESC también debe justificarse y equilibrarse a la luz de estos dos principios.

En este sentido, hay líneas rojas importantes que las instituciones estatales deben observar al momento de evaluar y decidir si la adopción de un sistema algorítmico es realmente apropiada a los fines pretendidos. Entre otros, la utilización de estos sistemas *no* se muestra adecuada *cuando*:

- la decisión debe basarse en la prudencia, el razonamiento o la experiencia humanos y el reconocimiento automático de patrones en los conjuntos de datos disponibles no tiene un papel útil para informar la intervención humana;
- el fin perseguido es difícil de traducir en variables matemáticas (ej., felicidad de la población) y no existen indicadores adecuados que puedan medirse a partir de los datos disponibles. Incluso si algunos *proxies* pudieran funcionar, la traducción matemática de un fenómeno social complejo y el uso de *proxies* relacionados puede no ser el enfoque más adecuado para abordar la cuestión o el problema;
- el estado actual de la tecnología (y del sistema específico que se considera adoptar) queda corto en atender las salvaguardias y derechos que deben aplicarse teniendo en cuenta la finalidad de la utilización del sistema (ej., la adopción de un sistema no explicable para apoyar la evaluación de riesgos por una autoridad administrativa o judicial);
- los desarrolladores y proveedores del sistema no ofrezcan garantías suficientes de que estos son fiables y seguros, alineados a estándares de derechos humanos que incluyen transparencia, no discriminación, privacidad, seguridad, entre otros atributos;
- el sistema presentar un historial de violaciones de derechos humanos;
- los desarrolladores y proveedores impidan que las instituciones públicas a cargo del despliegue del sistema y auditores independientes tengan acceso al código fuente y a los ejecutables del sistema, a los conjuntos de datos de entrenamiento anonimizados y a los materiales de prueba, incluidos los conjuntos de datos de prueba anonimizados;
- el único propósito de la adopción del sistema es reducir costes. El fin legítimo perseguido por la adopción del sistema debe estar relacionado con la protección y/o promoción de derechos humanos más allá de la mera reducción de costes.

Los Estados deben abstenerse de adoptar decisiones basadas en sistemas algorítmicos en contextos que serían incompatibles con los derechos humanos. Es decir, finalidades, contextos o sistemas que:

- se basen en prácticas del Estado que reproduzcan la discriminación sistémica y/o conlleven a la elaboración de perfiles raciales (ej., la aplicación de sistemas de reconocimiento facial para la seguridad pública y los sistemas de predicción de delitos);
- tengan impacto desproporcionado en grupos específicos y/o intrínsecamente reproduzcan visiones o prácticas discriminatorias reflejadas en conjuntos de datos sesgados utilizados para entrenar el modelo de IA o alimentar el funcionamiento del sistema o que, de otro modo, resulta de o reproduce perjuicios o estereotipos (los ejemplos señalados en el punto anterior, así como sistemas/ políticas de protección social que señalan a grupos específicos como potenciales defraudadores);
- se configuren como vigilancia indiscriminada de espacios físicos o virtuales;
- impliquen una regresión en garantías de protección social;
- representen la “instrumentalización” de los individuos (ej. explotar los datos de los más vulnerables o probar soluciones experimentales en poblaciones marginadas como incentivos para la colaboración con el sector privado).

D. Marcos regulatorios e institucionales y el impulso de nuevos desarrollos normativos.

Unos marcos de contratación pública eficaces son vitales para garantizar el uso responsable de la IA, fomentar la confianza, configurar el desarrollo ético de la IA y garantizar el aumento del valor público neto.

Ejemplos a escala internacional:

A nivel internacional, la Resolución A/HRC/50/56 de la ONU proporciona directrices para la adquisición, compra y concesión de licencias de IA para la prestación de servicios públicos.³³

Las Recomendaciones 2021 de la UNESCO sobre la Ética de la Inteligencia Artificial representan el primer acuerdo global que esboza valores y principios compartidos para guiar el desarrollo de marcos legales para el avance ético y sostenible de la IA que se alinea con los derechos humanos y los ODS.³⁴

El documento *AI Procurement in a Box* del Foro Económico Mundial expone las consideraciones fundamentales que deben tener en cuenta los gobiernos antes de adquirir y desplegar soluciones de IA, incluida la importancia de las evaluaciones iniciales del impacto

³³ <https://documents.un.org/doc/undoc/gen/g22/323/96/pdf/g2232396.pdf>

³⁴ <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

de la IA, la sensibilidad de los datos, la calidad de los datos y el consentimiento. También considera el impacto socioeconómico, especialmente en las poblaciones vulnerables.³⁵

El informe de la Asamblea General de las Naciones Unidas del Relator Especial (A/79/520) sobre el derecho a la educación destaca cómo la IA pone en tela de juicio los pilares fundamentales de la educación -currículo, pedagogía y evaluación- y propone un enfoque basado en los derechos humanos que haga hincapié en la participación, la rendición de cuentas, la no discriminación y la transparencia para orientar la integración de la IA en la educación.³⁶

También es importante remitirse a los principios y marcos sectoriales existentes, como las normas medioambientales y sanitarias, para defender los principios rectores establecidos sobre distintos aspectos que ahora se ven afectados por la IA. Por ejemplo, el Convenio de Aarhus de la CEPE³⁷ otorga a los ciudadanos derechos de información y participación en asuntos medioambientales, mientras que el Protocolo de Nagoya garantiza el acceso equitativo y la participación en los beneficios derivados de la utilización de los recursos genéticos, lo que adquiere relevancia dada la creciente integración de la IA en la agricultura, los sistemas alimentarios y la secuenciación genética.³⁸

Ejemplo a nivel nacional

La Ley de IA de la UE es una legislación histórica que adopta un enfoque basado en el riesgo, prohibiendo las aplicaciones de IA de alto riesgo, como los sistemas manipuladores o explotadores, al tiempo que impone estrictos requisitos de gobernanza, gestión de riesgos y transparencia.³⁹ El Proyecto de Declaración de Derechos de la IA de EE.UU. proporciona un marco importante para que el gobierno, las empresas tecnológicas y los ciudadanos puedan trabajar juntos para garantizar una IA más responsable.⁴⁰

En octubre de 2024, la Casa Blanca publicó un memorando sobre el avance de la adquisición responsable de IA en el gobierno, en el que se esbozan nuevos requisitos de contratación para las agencias federales.⁴¹

35 https://www3.weforum.org/docs/WEF_AI_Procurement_in_a_Box_AI_Government_Procurement_Guidelines_2020.pdf

36 <https://www.ohchr.org/en/documents/thematic-reports/a79520-artificial-intelligence-education-report-special-rapporteur-right>

37 <https://unece.org/environment-policy/public-participation/aarhus-convention/introduction>

38 <https://www.cbd.int/abs/default.shtml>

39 <https://artificialintelligenceact.eu/the-act/>

40 <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>

41 <https://www.cov.com/en/news-and-insights/insights/2024/10/omb-releases-requirements-for-responsible-ai-procurement-by-federal-agencies>

En Nueva Zelanda, el Ministro de Estadística lanzó la Carta de Algoritmos, comprometiendo a los organismos gubernamentales a mejorar la transparencia y la rendición de cuentas en su uso de algoritmos.⁴² La ley brasileña sobre IA otorga a las personas y grupos afectados por sistemas de IA de alto riesgo derechos fundamentales, como el derecho a una explicación de las decisiones del sistema, el derecho a impugnar y solicitar revisiones, y el derecho a la supervisión humana de las decisiones.

E. Recomendaciones generales:

a) Sobre la moderación automatizada de contenido. Para mitigar los impactos negativos de la moderación automatizada, el informe propone varias estrategias:

1. Mayor transparencia y rendición de cuentas

- Las plataformas deben publicar de manera clara y accesible los criterios utilizados para la moderación de contenido.
- Se alienta que las plataformas realicen auditoría externa de los sistemas automatizados de moderación para evaluar su impacto en derechos humanos.

2. Implementación de mecanismos efectivos de apelación

- Los usuarios deben tener derecho a impugnar decisiones de eliminación de contenido.
- Las plataformas deben contar con revisiones humanas en casos donde se afecte la libertad de expresión.

3. Regulación basada en derechos humanos

- Es necesario un marco normativo que garantice que la moderación de contenido respete la libertad de expresión y otros derechos fundamentales.
- Los gobiernos deben desarrollar estándares en línea con el marco interamericano de derechos humanos.

⁴² https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf

4. Inclusión de perspectivas diversas en el diseño de algoritmos y mejorar la moderación automatizada.

- Es fundamental la participación de comunidades vulnerables en la creación de políticas de moderación de contenido.
- Es necesario que los algoritmos consideren en el contexto cultural y lingüístico para evitar la eliminación injusta de contenido. Esto trae a colación que aquellos mecanismos ya creados por las plataformas digitales sobre revisión humana deben fortalecerse, ser transparentes y sobre todo contar con la posibilidad de apelación de manera efectiva, sencilla y rápida.

b) Sobre los derechos humanos relacionados con la adquisición y el despliegue de IA por el poder público

La importancia de una estructura subyacente de protección de derechos en atención a los principios transversales presentados anteriormente

Consideramos que existe un conjunto de derechos que representan principios que los Estados deben observar desde la evaluación inicial sobre la adopción de sistemas algorítmicos para fines que afectan derechos, hasta la etapa de monitoreo y evaluación de estas políticas y proyectos. Estos principios transversales corresponden a un aparato basilar que los Estados deben tener establecido, o contar con medidas efectivas y avanzadas en la dirección de establecerlos, para el despliegue adecuado de dichos sistemas. Contar con este aparato no es secundario; se deriva de la obligación de los Estados de prevenir violaciones de derechos humanos y de su papel fundamental como garantes de estos derechos. Así nuestras recomendaciones involucran desarrollos concretos de los principios transversales presentados anteriormente de manera a asegurar la existencia y debida operación de este aparato basilar. Los presentamos a continuación:

Participación social: una participación social significativa requiere funcionarios estatales dedicados, presupuesto, procesos y planificación, para que no sea un mero ejercicio de cumplimiento formal sin sustancia real.

Igualdad y no discriminación: cumplir con estos principios también implica contar con la estructura necesaria para una participación significativa, especialmente para involucrar a grupos que han sido históricamente discriminados. Requiere movilizar diversos conocimientos expertos dentro y fuera de las instituciones estatales para evaluar adecuadamente el sistema y el contexto social en el que se aplicará, y abordar posibles

problemas. Exige un monitoreo continuo de la implementación, con supervisión humana competente y responsable. Finalmente, implica la producción y análisis de indicadores orientados a los derechos humanos sobre la implementación del sistema y sus impactos positivos y negativos.

Debido proceso/acceso a la justicia: este principio demanda recursos judiciales y administrativos fácilmente accesibles, equitativos y efectivos, así como estructuras adecuadas para investigar y sancionar violaciones de derechos humanos resultantes del uso estatal de sistemas algorítmicos, garantizando reparación y no repetición. No es raro que la implementación de sistemas algorítmicos por parte de los Estados afecte directamente el reconocimiento y ejercicio de derechos esenciales sin proporcionar mecanismos de apelación eficientes y accesibles.

Transparencia/acceso a la información: requiere rutinas y personal para producir, organizar y divulgar información, tanto de manera proactiva como en respuesta a solicitudes, así como un organismo de supervisión independiente con poderes suficientes y efectivos.

Privacidad y protección de datos: estos principios exigen que las instituciones estatales que evalúan e implementan sistemas de IA/ADM cuenten con una infraestructura de seguridad y experiencia adecuadas. La implementación de proyectos basados en algoritmos que involucren el procesamiento de datos personales también debe estar bajo la supervisión de un organismo independiente de protección de datos y de departamentos estatales o funcionarios dentro de las instituciones responsables del programa que puedan desempeñar el papel de oficiales de protección de datos (ante los cuales las personas puedan ejercer sus derechos relacionados con los datos, como acceso y rectificación). Asimismo, deben existir medidas y procedimientos para satisfacer oportunamente los derechos de los titulares de los datos.

Más sobre la participación cívica y la transparencia

La participación cívica y el compromiso público son esenciales para diseñar y aplicar políticas de responsabilidad algorítmica. Para garantizar una participación pública significativa, pueden tenerse en cuenta las siguientes recomendaciones:

- Llevar a cabo consultas eficaces de interés público en las que participen múltiples grupos de interés. Para fomentar la equidad y la inclusión, considere la posibilidad de emplear métodos como el sondeo deliberativo, los paneles de referencia de ciudadanos, los jurados de ciudadanos y los presupuestos participativos.⁴³ Un ejemplo notable es la Ordenanza de Vigilancia y Seguridad Comunitaria de Oakland, que exige amplias audiencias públicas a través de foros establecidos como las reuniones del consejo.⁴⁴

⁴³ <https://www.oecd.org/gov/open-government/eight-ways-to-institutionalise-deliberative-democracy.htm>

⁴⁴ <https://www.opengovpartnership.org/wp-content/uploads/2021/08/algorithmic-accountability-public-sector.pdf>

- Colaborar con las comunidades marginadas y las organizaciones que representan a grupos marginados para comprender y eliminar las barreras específicas a las que se enfrentan. La equidad y la accesibilidad deben garantizarse eliminando las barreras a su participación, proporcionando ajustes razonables (por ejemplo, intérpretes de lengua de signos, intérpretes de lenguas indígenas), compensando a los representantes, utilizando formatos y canales de comunicación accesibles, y financiando y apoyando programas de desarrollo de capacidades.⁴⁵
- Basar los procesos de toma de decisiones relacionados con la IA en marcos establecidos de derechos de participación ex ante, como el Convenio de Aarhus de la CEPE.⁴⁶
- Exigir la transparencia de los sistemas de IA y de sus procesos de adquisición y despliegue.⁴⁷ Muchos países carecen de transparencia a la hora de adquirir y desplegar tecnologías de IA, con procesos de contratación opacos que alimentan la corrupción y limitan la competencia.⁴⁸ Los organismos públicos deben ser conscientes del tipo de información que comparten y de cómo determinados públicos acceden a ella, confían en ella o la utilizan. Por ejemplo, los registros de algoritmos de Ámsterdam y Helsinki se diseñaron específicamente para audiencias críticas como la sociedad civil, mientras que Francia emplea explicaciones audiovisuales en lenguaje sencillo para informar a las comunidades afectadas y al público.⁴⁹

Interesa recordar que los mecanismos participativos interamericanos en el contexto de las comunidades indígenas y afrodescendientes aportan valiosos modelos y lecciones para impulsar una participación significativa también en otros contextos – especialmente respecto a lo que se entiende por consulta previa, libre e informada a comunidades afectadas. Algunos de ellos son:

- La consulta no es un acto singular, sino un proceso de diálogo en el que se facilita información clara, accesible y completa con tiempo suficiente para permitir un compromiso adecuado⁵⁰;

⁴⁵ https://gpai.ai/projects/responsible-ai/towardsrealdiversityandgenderequalityinai/towards-substantive-equality%20in-artificial-intelligence_Transformative-AI-policy-for-gender-equality-and-diversity.pdf

⁴⁶ <https://unece.org/environment-policy/public-participation/aarhus-convention/introduction>.

⁴⁷ https://ia.derechosdigitales.org/wp-content/uploads/2024/12/2024-LATAM-IA_en_el-Estado-ES.pdf

⁴⁸ https://ia.derechosdigitales.org/wp-content/uploads/2024/12/2024-LATAM-IA_en_el-Estado-ES.pdf; https://www.oecd.org/content/dam/oecd/en/publications/reports/2021/07/public-procurement-in-the-state-of-mexico_df343ad7/cc1da607-en.pdf

⁴⁹ <https://www.opengovpartnership.org/wp-content/uploads/2021/08/executive-summary-algorithmic-accountability.pdf>

⁵⁰ CIDH, Derechos de los Pueblos Indígenas y Tribales sobre sus Tierras Ancestrales y Recursos Naturales: Normas y Jurisprudencia del Sistema Interamericano de Derechos Humanos 30 de diciembre de 2009, párr. 285, y CIDH, Pueblos Indígenas, Comunidades

- La consulta de buena fe requiere la ausencia de cualquier tipo de coacción y debe ir más allá de los procedimientos meramente pro forma⁵¹;
- No tener debidamente en cuenta los resultados de la consulta es contrario al principio de buena fe⁵²;
- Las decisiones resultantes del proceso de consulta están sujetas a las autoridades administrativas y judiciales superiores, a través de procedimientos adecuados y eficaces, para evaluar su validez, pertinencia y el equilibrio entre los derechos e intereses en juego⁵³.

Como mencionado, la transparencia es elemento clave para garantizar el derecho de acceso la información, la participación cívica significativa, entre otros derechos directamente involucrados en el uso estatal de sistemas algorítmicos. El informe de la EFF sobre la aplicación de estándares interamericanos de DDHH en este contexto desarrolla implicaciones y deberes específicos [en materia de transparencia](#) que se pueden consultar [a más detalle](#). Las recomendaciones generales incluyen:

- Dar prioridad al desarrollo, adquisición e implantación de sistemas algoritmos de código abierto. Poner a disposición del público el código fuente del sistema es una medida de transparencia crucial, que instamos a los Estados a adoptar para los sistemas que apoyen decisiones consecuenciales sobre la vida de las personas. En cualquier caso, una evaluación adecuada del sistema antes y durante su implantación implica que la institución estatal responsable por la implementación y expertos independientes tengan acceso y analicen su código fuente;
- Poner a disposición del público la documentación de las especificaciones del sistema, estableciendo esto como requisito previo para que los desarrolladores y proveedores puedan contratar con las instituciones estatales;
- Adoptar únicamente modelos que sean interpretables⁵⁴ o que sigan un planteamiento de explicabilidad mediante el diseño⁵⁵.

Afrodescendientes y Recursos Naturales: Protección de Derechos Humanos en el Contexto de Actividades de Extracción, Explotación y Desarrollo, 31 de diciembre de 2015, párr. 108.

⁵¹ CIDH, Derechos de los Pueblos Indígenas y Tribales sobre sus Tierras Ancestrales y Recursos Naturales., párr. 318.

⁵² CIDH, Derechos de los Pueblos Indígenas y Tribales sobre sus Tierras Ancestrales y Recursos Naturales., párr. 325.

⁵³ CIDH, Derechos de los Pueblos Indígenas y Tribales sobre sus Tierras Ancestrales y Recursos Naturales., párr. 328.

⁵⁴ Rudin, C. (2019). Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat Mach Intell* 1, 206-215. See also Hamon, R., Junklewitz, H., & Sanchez, I. (2020), p. 13 (sección 3.2.4 Interpretable models vs. post-hoc interpretability).

⁵⁵ Hamon, R., Junklewitz, H., & Sanchez, I. (2020), p. 24.

c) Sobre el uso de IA en procuración de justicia.

1. Capacitación y habilidades digitales:

- Diseñar programas de formación continua en IA, análisis de datos y ciberseguridad.
- Crear equipos multidisciplinarios (juristas + tecnólogos) para desarrollar herramientas alineadas con derechos humanos.

2. Transparencia y rendición de cuentas:

- Publicar normativas, avisos de privacidad y protocolos de IA en portales de transparencia.
- Establecer mecanismos claros para auditar algoritmos y evitar sesgos.

3. Protección a derechos humanos, incluyendo la protección de datos:

- Realizar evaluaciones de impacto a los derechos humanos antes de implementar IA.
- Garantizar el consentimiento informado y los derechos ARCO en el tratamiento de datos.

4. Ciberseguridad y gestión documental:

- Adoptar estándares [ISO 27001](https://www.iso.org/es/norma/27001)⁵⁶ para proteger información sensible.
- Digitalizar expedientes físicos y crear protocolos de respaldo seguro.

⁵⁶ <https://www.iso.org/es/norma/27001>