elDial.com
Contenidos Jurídicos

Los riesgos de la inteligencia artificial para el principio de igualdad y no discriminación.
Planteo de la problemática y algunas aclaraciones conceptuales necesarias bajo el prisma del Sistema Interamericano de Derechos Humanos.

Por Mariana Sánchez Caparrós[(*)]

Sumario: 1. Introducción. 2. Inteligencia artificial como sistema sociotécnico. 2.1. ¿Qué es la inteligencia artificial? 2.2. La inteligencia artificial como sistema sociotécnico. 3. Los riesgos de la inteligencia artificial para el derecho a la igualdad y no discriminación. 3.1. La inteligencia artificial y sus riesgos asociados. 3.2. Desafíos para el principio de igualdad y no discriminación. 3.3. La problemática de los *proxies*. 4. Sesgo y discriminación algorítmica: aclaraciones conceptuales relevantes desde la perspectiva del Sistema Interamericano de Derechos Humanos. 4.1. ¿Qué es el sesgo algorítmico? 4.2. El principio de igualdad y no discriminación en el Sistema Interamericano de Derechos Humanos. 4.3. Recapitulación. 5. Conclusiones.

1. Introducción.

El empleo de sistemas inteligentes para la toma de decisiones plantea nuevos desafíos para el derecho. Uno de ellos, sin dudas, es el relacionado con los riesgos que trae para el principio de igualdad y no discriminación, en tanto estos sistemas pueden contribuir a reproducir y perpetuar estereotipos y prejuicios sociales que están presentes en los datos de entrenamiento o que provienen de las personas que intervienen a lo largo de su ciclo de vida, provocando situaciones de discriminación[1].

Hace sólo un par de años, el *chatbot* Tay, desarrollado por Microsoft Corporation, que había sido diseñado para mantener conversaciones informales y divertidas con los usuarios en la red social Twitter, a sólo 24 horas de haber sido colocado en internet tuvo que ser desactivado por la empresa debido a que comenzó a emitir comentarios e insultos racistas, xenófobos y sexistas hacia otros usuarios[2].

[(*)] Relatora en el Superior Tribunal de Justicia de la Provincia de Tierra del Fuego AeIAS. Abogada (UBA, Diploma de honor), Mg. en Derecho Administrativo (Universidad Austral, Diploma de honor y tesis premiada) y Doctorando en Cs. Jurídicas (UCA). Autora del libro *Categorías Sospechosas* (Astrea, 2020) y coautora del *Tratado de Inteligencia Artificial y Derecho* (La Ley, 2021). Profesora de grado (UCES y UNTDF). Profesora del Programa Blockchain, Derecho y Empresa (Universidad Austral) y profesora invitada del Postgrado en Inteligencia Artificial y Derecho (UBA-IALAB), así como de sus cápsulas en IA y Administración Pública y IA y Género. Conferencista. Autora de artículos de doctrina de la especialidad derecho público y derecho y tecnología. Miembro del comité editorial de la Revista en Legaltech y Derecho 4.0 de IJ Editores y de Legaltech Seed.

[1] COMISIÓN EUROPEA, *Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza*, Bruselas, 2020, p. 14, en https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificialintelligence-feb2020_es.pdf [accedido el 5/5/2022].

[2] BBC MUNDO, "Tay, la robot racista y xenófoba de Microsoft", 25 de marzo de 2016, en https://www.bbc.com/mundo/noticias/2016/03/160325_tecnologia_microsoft_tay_bot_adolescente_inteligencia_artificial_racista_xenofoba_lb [accedido el 5/5/2022].

Un esfuerzo coordinado de algunos usuarios de la red social había logrado, en muy poco tiempo, que este *bot* de perfil adolescente y rebelde adoptara una personalidad horrible, que no sólo declaraba abiertamente su empatía hacia Hitler y su apoyo al genocidio, sino que también vertía insultos aberrantes hacia otras personas[3].

En el año 2015, Google debió pedir disculpas públicas luego de que su aplicación para compartir y guardar imágenes y videos —*Google Photos*—, que cuenta con un sistema inteligente de etiquetado, identificara por error a una pareja de color como "gorilas"[4].

Y más cerca en el tiempo, en septiembre del año 2021, algo similar sucedió con la empresa Facebook —ahora Meta—, que debió disculparse públicamente luego de que su sistema inteligente de recomendaciones colocara un mensaje automático preguntando a los usuarios si querían seguir viendo videos sobre "primates" tras la proyección de uno en el que aparecían hombres de color [5].

Además de los descriptos, existen muchos otros ejemplos en los que se pueden observar cómo los sistemas inteligentes, una vez desplegados en el mundo real, provocan discriminación en las personas o contribuyen a perpetuar la situación de vulnerabilidad de ciertos colectivos de la sociedad[6].

El empleo cada vez mayor de sistemas inteligentes en la sociedad, plantea numerosos desafíos en materia de protección de derechos fundamentales y, en particular, del derecho a la igualdad y no discriminación, que es uno de los pilares del sistema democrático.

El presente trabajo se propone indagar brevemente en la problemática existente en torno al sesgo de los algoritmos que puede derivar en situaciones de discriminación, para entonces efectuar algunas aclaraciones respecto del alcance de los términos "sesgo algorítmico" y "discriminación algorítmica", usualmente empleados cuando se aborda la cuestión de la discriminación algorítmica, a los fines de sentar algunas bases para su adecuada distinción y empleo.

A ese fin, se comenzará por repasar brevemente qué es la inteligencia artificial y cuáles son los riesgos que trae asociados el empleo de estos sistemas, para entonces examinar cuál es aquél que se presenta, en particular, con relación al principio de igualdad y no discriminación.

Finalmente, el trabajo concluye con un examen de la normativa protectoria del principio de igualdad y no discriminación contenida en la Convención Americana de Derechos Humanos, para a partir de ella intentar echar luz sobre los conceptos de "sesgo algorítmico" y "discriminación algorítmica", en pos de determinar su adecuado alcance y también identificar si permiten comprender adecuadamente el fenómeno de la discriminación cuando se produce como consecuencia del empleo de sistemas inteligentes.

---

[3] Sobre el fracaso de Tay ampliar en Rachel METZ, "La IA racista y neonazi de Microsoft fue culpa de Microsoft", *MIT Technology Review*, marzo 2016, en https://www.technologyreview.es/s/5720/la-ia-racista-y-neonazi-de-microsoft-fue-culpa-de-microsoft [accedido el 5/5/2022]. Asimismo ver: BBC MUNDO, "Tay, la robot racista…".

[4] BBC MUNDO, "Google pide perdón por confundir a una pareja negra con gorilas", 15 de julio de 2015, https://www.bbc.com/mundo/noticias/2015/07/150702_tecnologia_google_perdon_confundir_afroamericanos_gorilas_lv [accedido el 24/2/2022].

[5] INFOBAE, "Facebook lamenta etiqueta en video de hombres de raza negra", 4 de septiembre de 2021, https://www.infobae.com/america/agencias/2021/09/05/facebook-lamenta-etiqueta-en-video-de-hombres-de-raza-negra/, accedido el 26/2/2022.

[6] En 2014 Amazon diseñó un sistema inteligente de asistencia en la búsqueda de personal que terminó discriminando a las candidatas mujeres y tuvo que ser abandonado. Ampliar en Jeffrey DASTIN, "Amazon scraps secret AI recruiting tool that showed bias against women", 10 de octubre de 2018, *Reuters*, disponible en www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-airecruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G [accedido el 25/2/2022].

2. Inteligencia artificial como sistema sociotécnico.

2.1. ¿Qué es la inteligencia artificial?

El término "inteligencia artificial" fue introducido en los años '50 para describir las máquinas que podían hacer más que tareas rutinarias[7]. En concreto, se trata de un término que fue acuñado en 1955 por un grupo de investigadores —John McCarthy, Marvin L. Minsky, Nathaniel Rochester and Claude E. Shannon— que organizó el evento que luego sería reconocido como el que dio inicio al estudio de la IA: el workshop de verano llamado "Estudio de la Inteligencia Artificial"[8].

Con el tiempo, este campo de investigación y desarrollo se amplía y diversifica, dando lugar a numerosas maneras de definir el término "inteligencia artificial". Ello al punto de que, al día de la fecha, no existe una única definición de inteligencia artificial que sea universalmente aceptada[9], sino diferentes definiciones que nacen de las distintas disciplinas que han abordado a la IA como objeto de estudio, tal como las ciencias de la computación, la ingeniería, la robótica, la psicología o la filosofía[10].

De hecho, en el documento titulado "Recomendación sobre la ética de la inteligencia artificial" producido por la UNESCO a fines del año 2021, el organismo internacional concluyó que para el abordaje de la cuestión ética, en lugar de lograr una definición común y en tanto se trata de una noción que tiende a modificarse con el tiempo en función de los avances tecnológicos que acontecen, resultaba más adecuado enfocarse en las características de los sistemas inteligentes que adquieren importancia cuando se estudia el tema[11].

En virtud de ello, para el propósito de este trabajo se tomará una definición amplia de IA, que la describe como un conjunto de técnicas y tecnologías guiadas por un grupo de objetivos definido por seres humanos, que cuentan con cierto grado de autonomía para tomar decisiones o hacer predicciones en base a los datos que recibe[12].

Por "sistema inteligente" se entenderá a todo sistema de *software* o *hardware* que incorpore IA[13]. En este trabajo, uno y otro término se emplearán de forma indistinta para referirse a la IA.

2.2. La inteligencia artificial como sistema sociotécnico.

---

[7] COMISIÓN MUNDIAL DE ÉTICA DEL CONOCIMIENTO CIENTÍFICO Y LA TECNOLOGÍA, "Preliminary Study on the Ethics of Artificial Intelligence", Paris, UNESCO, 26 de febrero de 2019, en https://unesdoc.unesco.org/ark:/48223/pf0000367823 [accedido el 4/4/2022], p. 3.

[8] Nick BOSTROM, "The Ethics of Artificial Intelligence", *Draft for Cambridge Handbook of Artificial Intelligence*, Londres, 2011, en https://www.nickbostrom.com/ethics/artificial-intelligence.pdf [accedido el 4/3/2022], p. 2. También: COMISIÓN MUNDIAL DE ÉTICA DEL CONOCIMIENTO CIENTÍFICO Y LA TECNOLOGÍA, "Preliminary Study on…", p. 5.

[9] Ampliar en Sofia SAMOILI, Montserrat LÓPEZ COBO, Blajob DELIPETREV, Fernando MARTÍNEZ-PLUMED, Emilia GÓMEZ y Giuditta DE PRATO, *AI Watch. Defining Artificial Intelligence 2.0: Towards an operational definition and taxonomy for the AI landscape*, Publications Office of the European Union, Luxembourg, 2021, en https://publications.jrc.ec.europa.eu/repository/handle/JRC118163 [accedido el 30/4/2022.]

[10] COMISIÓN MUNDIAL DE ÉTICA DEL CONOCIMIENTO CIENTÍFICO Y LA TECNOLOGÍA, "Preliminary Study on…" p. 5.

[11] UNESCO, "Recomendación sobre la ética de la inteligencia artificial", 22 de noviembre de 2022, en https://unesdoc.unesco.org/ark:/48223/pf0000379920_spa.page=15 [accedido el 3/3/2022], p. 16.

[12] Axel VOSS, "Draft Report on artificial intelligence in a digital age", *European Parliament*, noviembre de 2021, p. 38, en https://www.europarl.europa.eu/doceo/document/AIDA-PR-680928_EN.pdf [accedido el 12/5/2022].

[13] GRUPO DE EXPERTOS DE ALTO NIVEL SOBRE IA, "Directrices éticas para una IA fiable", abril de 2019, en https://urjes.com/pruebas/wp-content/uploads/2021/05/Grupo-independiente-de-expertos-...-Directrices-eticas-para-una-inteligencia-artificial-IA-fiable-ES.pdf [accedido el 26/2/2022], p. 48.

Los sistemas inteligentes integran modelos y algoritmos que, a partir de los datos disponibles y la capacidad de cómputo, pueden aprender y realizar tareas cognitivas que dan lugar "…a resultados como la predicción y la adopción de decisiones en entornos materiales y virtuales…"[14].

Respecto de las diversas técnicas de IA, se ha señalado, "...se basan en detectar y reconocer patrones de información en los datos (…) a partir de combinar ordenadores, internet, algoritmos[15] y lenguajes de programación para resolver problemas o tomar decisiones que antes solo podían ser realizadas por nuestras capacidades cognitivas..."[16].

Es importante tener presente que la IA que conocemos hasta el momento pertenece al ámbito de la IA restringida, que comprende aquellos sistemas informáticos que sólo pueden ejecutar tareas programadas y brindar soluciones en dominios particulares del conocimiento humano[17].

Como contrapartida, la IA general —cuyos desafíos éticos no serán abordados en este trabajo— describe sistemas que aún no existen, que serían capaces de reproducir nuestra habilidad para gestionar varios campos de conocimiento a la vez, así como nuestro sentido común. Es decir, sistemas que simulan enteramente el comportamiento humano[18].

En cualquier caso, y más allá de cómo se defina a la IA, así como de sus alcances —IA general vs. IA restringida—, los sistemas inteligentes deben ser comprendidos como sistemas sociotécnicos, es decir, sistemas cuyo objeto de transformación lo constituyen los seres humanos[19].

Este enfoque permite considerar que se trata de sistemas más complejos que las construcciones matemáticas y computacionales que les dan sostén, para poder observarlos de manera integral y, desde esa óptica, estudiar la interacción que se produce entre las personas y la IA.
Esto es, cómo las primeras influyen en lo relativo a su desarrollo desde una mirada amplia, y también cómo ésta última impacta para transformar y reconfigurar la vida individual y social de los seres humanos.

Se trata de una visión que permite tener en cuenta, por su relevancia, los valores, sesgos y comportamientos que vienen embebidos en los sets de datos y en los humanos que intervienen en las diversas etapas del ciclo de vida del sistema, así como otros factores propios de la complejidad de las organizaciones en las que éstos se crean y que tienen impacto en su diseño, desarrollo y despliegue.

Desde dicha perspectiva, las consideraciones acerca del impacto de la IA atenderán, no sólo de cómo se diseña el sistema desde lo eminentemente técnico -v.gr. que algoritmo o modelo se selecciona-, sino también a la manera en que éste se desarrolla y despliega en un contexto más amplio, lo que alcanza tanto los datos utilizados, el propósito declarado del sistema, su

---

[14] UNESCO, "Recomendación sobre…", p. 16.
[15] Un algoritmo es una estructura de control finita, abstracta, efectiva y compuesta, dada de manera imperativa, que cumple un propósito dado bajo provisiones dadas. Cfr. Robin K. HILL, "What an Algorithm Is", *Philos. Technol*, 11 de mayo de 2015, p. 47, en https://link.springer.com/article/10.1007/s13347-014-0184-5 [accedido el 6/4/2022].
[16] Juan G. CORVALÁN, *Perfiles Digitales Humanos*, La Ley, Buenos Aires, 2020, p. 33.
[17] ASAMBLEA GENERAL DE LAS NACIONES UNIDAS, p. 4, "Resolución 73/348", en http://undocs.org/es/A/73/348 [accedido el 15/3/2022].
[18] Juan G. CORVALÁN, *Perfiles Digitales*…, p. 37/39.
[19] Los sistemas sociotécnicos son "…aquellos sistemas técnicos cuyo objeto de transformación lo constituyen los propios seres humanos (…) las sociotecnologías forman parte de la rama general de la tecnología, siendo su objeto el diseño de planes de acción para transformar lo social y para resolver los problemas que surgen en las sociedades…". Cfr. German HEVIA MARTÍNEZ, "La sociedad como artefacto. Sistemas sociotécnicos, sociotecnologías y sociotécnicas", *Revista Iberoamericana de Ciencia Tecnología y Sociedad* 40, febrero de 2019, p. 267, en https://www.researchgate.net/publication/332863180_La_sociedad_como_artefacto_Sistemas_sociotecnicos_sociotecnologias_y_sociotecnicas [accedido el 15/5/2022].

funcionalidad y precisión, la escala de implementación y el contexto organizacional, social y legal en el que se lo implementa[20].

3. Los riesgos de la inteligencia artificial para el principio de igualdad y no discriminación.

3.1. La inteligencia artificial y sus riesgos asociados.

El mundo se enfrenta a un crecimiento acelerado del uso de sistemas con capacidad para aprender y realizar tareas cognitivas que antes sólo podían ser realizadas por las personas[21]. Y cada vez es más consciente de que el potencial de los sistemas inteligentes para mejorar el bienestar social e individual[22], viene acompañado de riesgos significativos para los derechos humanos.

De hecho, hay quienes han llegado a señalar que lo que se pone en riesgo frente a la posibilidad de un desarrollo de la IA sin ningún tipo de regulación ética ni jurídica es la supervivencia misma de la especie humana, pues estos sistemas tienen el potencial de afectar gravemente nuestros derechos[23].

El poder creativo y destructivo de la inteligencia artificial, se ha apuntado, es comparable con el de la energía nuclear, "…en el sentido de que ambas tecnologías, bajo ciertas variables, están en condiciones de provocarle un daño devastador a la especie humana…"[24].

Se ha apuntado que estos sistemas representan un riesgo para la autonomía de las personas debido a su capacidad de influir en su comportamiento, ya que a partir de los datos que emplean como combustible pueden perfilarlas[25] para darles "pequeños empujones"[26] que inciden en sus decisiones sin que sean conscientes de ello[27].

---

[20] COUNCIL OF EUROPE, *A legal framework for AI systems. Feasibility study for a legal framework for the development, design and application of artificial intelligence, based on Council of Europe´s standards on human rights, democracy and the rule of law*, adoptado por la CAHAI en su 3ra reunión plenaria el 17 de diciembre de 2020, p. 7, en https://media-exp1.licdn.com/dms/document/C4D1FAQEe4eGCLGf6CA/feedshare-document-pdf-analyzed/0/1649063227482?e=2147483647&v=beta&t=ODWWaRAOoF3tufI-aNt1wCP57PWKqf8mk8FkJ6alA58 [accedido el 5/5/2022].

[21] COMISIÓN MUNDIAL DE ÉTICA DEL CONOCIMIENTO CIENTÍFICO Y LA TECNOLOGÍA, "Preliminary Study …", p. 3.

[22] A modo de ejemplo, el dispositivo *Or Cam* que se acopla a cualquier lente, se activa por comando de voz y permite leer instantáneamente el texto de libros, pantallas y otras superficies, reconocer rostros e identificar el lugar en que están ubicados los objetos, ayudando a las personas con discapacidad visual a vivir una vida más independiente. Ver https://www.youtube.com/watch?v=mwXGivl77Tg, [accedido el 22/5/2022].

[23] Cfr. Marcos LÓPEZ ONETO, "Fundamentos antropológicos, éticos, filosóficos, históricos, sociológicos y jurídicos para la constitución universal de un derecho de la inteligencia artificial (DIA)", en Juan G. CORVALÁN (Dir.), *Tratado de Inteligencia Artificial y Derecho*, La Ley, Buenos Aires, 2021, t. 1, p. 75 y 78.

[24] *Ibid.*, p. 75.

[25] El perfilado se refiere al proceso de emplear el reconocimiento de patrones y correlaciones para crear perfiles de usuarios que representan o identifican personas. Ampliar en Philip JANSEN y Philip BREY, "Ethical Analysis of AI and Robotics Technologies", *SIENNA Project*, 31 de agosto de 2019, p. 67, en https://www.sienna-project.eu/digitalAssets/801/c_801912-l_1-k_d4.4_ethical-analysis--ai-and-r--with-acknowledgements.pdf [accedido el 22/5/2022].

[26] La teoría del empujón, propuesta por Richard Thaler, parte de la premisa de que, entre dos opciones, las personas suelen escoger aquella que es más fácil sobre la que resulta más adecuada. Ampliar en BBC MUNDO, "Qué es la 'teoría del empujón' que le hizo ganar al estadounidense Richard H. Thaler el Nobel de Economía", 9 de octubre de 2017, en https://www.bbc.com/mundo/noticias-41551856 [accedido el 30/5/2022].

[27] David LESLIE, "Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector", *The Alan Turing Institute*, 2019, p.

También significan un riesgo a la privacidad desde que permiten la identificación, el monitoreo y seguimiento de las actividades y comportamientos de las personas, tanto en el espacio público como privado[28], no sólo vía el reconocimiento facial, sino también de voz y el *tracking* de sus movimientos a través de los dispositivos inteligentes que llevan consigo.

En suma, los sistemas inteligentes plantean nuevos y numerosos desafíos éticos desde que pueden afectar la autonomía personal para la toma de decisiones[29]; el futuro del empleo[30]; el acceso a la información[31], a la educación[32] y a la salud[33]; la democracia y el estado de derecho[34]; la brecha digital; la protección del consumidor y de los datos personales; el derecho a la privacidad y el derecho a la igualdad y a la no discriminación[35].

Todo ello en un contexto al que se agrega que ciertos enfoques de aprendizaje automático se caracterizan porque los pasos a través de los cuales el sistema ha arribado a una decisión se presentan inteligibles para los expertos humanos, incluidos, quienes han diseñado el sistema[36].

Esto último da lugar al denominado "problema de las cajas negras", noción que se emplea para describir la dificultad que sobreviene para explicar el resultado obtenido por la IA en términos comprensibles por el común de las personas, más cuyo abordaje excedería el objeto de estudio aquí planteado.

No obstante, sí puede señalarse que frente a los riesgos que suponen los sistemas inteligentes que se han mencionado, no quedan dudas de que la falta de transparencia acerca de cómo operan y las dificultades para lograr una plena explicabilidad, se convierten en un obstáculo adicional para su control.

3.2. Desafíos para el principio de igualdad y no discriminación en la era de la inteligencia artificial.

5, en https://www.turing.ac.uk/research/publications/understanding-artificial-intelligence-ethics-and-safety [accedido el 15/3/2022].

[28] *RETINA Tendencias*, "¿Qué es y en qué consiste el sistema de crédito social chino?", 5 de febrero de 2021, en https://www.youtube.com/watch?v=lun7V7TPmmM [accedido el 10/4/2022].

[29] Ricardo SENRA, "El dilema de las redes sociales' de Netflix: 5 secretos de los dueños de las redes para engancharnos y manipularnos, según el documental", *BBC News Brasil*, 3 de octubre de 2020, en https://www.bbc.com/mundo/noticias-54385775 [accedido el 10/4/2022].

[30] Laura RIPANI, Adriana KUGLER, Nicolás SOLER, Maurice KUGLER y Ramiro RODRIGO, *El futuro del trabajo en América Latina y el Caribe: ¿Cuál es el impacto de la automatización en el empleo y los salarios?*, Banco Interamericano de Desarrollo, diciembre de 2020, en https://publications.iadb.org/es/el-futuro-del-trabajo-en-america-latina-y-el-caribe-cual-es-el-impacto-de-la-automatizacion-en-el [accedido el 10/4/2022].

[31] Joseph PALERMO, "Deepfakes: Why you can´t believe everything you see and hear", 12 de febrero de 2020, e n https://www.youtube.com/watch?v=JXBBalHl_cI [accedido el 10/4/2022].

[32] Patricia TUBELLA, "Los estudiantes británicos derrotan al algoritmo de Johnson", *El País*, 18 de agosto de 2020, en https://elpais.com/internacional/2020-08-18/los-estudiantes-britanicos-propician-la-primera-gran-derrota-de-johnson.html [accedido el 10/4/2022].

[33] Javier SALAS, "El algoritmo que discrimina a los pacientes negros sin conocer su raza", *El País*, 24 de octubre de 2019, en https://elpais.com/elpais/2019/10/24/ciencia/1571909798_596622.html [accedido el 10/4/2022].

[34] Basta recordar el caso de Cambridge Analytica y Facebook. Ampliar en BBC MUNDO, "5 claves para entender el escándalo de Cambridge Analytica que hizo que Facebook perdiera U$S 37.000 millones en un día", 20 de marzo de 2018, en https://www.bbc.com/mundo/noticias-43472797 [accedido el 8/4/2022].

[35] UNESCO, "Recomendación sobre la ética…", p. 17.

[36] Marcelo IENCA y Effy VAYENA, "AI Ethics Guidelines: European and Global Perspectives", en CAHAI Secretariat (Comp.), *Towards Regulation of AI Systems*, diciembre de 2020, p. 51, en https://www.coe.int/en/web/artificial-intelligence/-/- toward-regulation-of-ai-systems- [accedido el 5/3/2022]. También David LESLIE, "Understanding artificial intelligence ethics and safety…".

Aun cuando los sistemas inteligentes se presentan a la comunidad bajo un ropaje de neutralidad, veracidad y cientificismo, lo cierto es que su naturaleza sociotécnica permite inferir que no son neutrales, porque no lo son los datos con los que se los entrena, ni tampoco las personas detrás de su diseño, definición de sus objetivos, desarrollo y despliegue[37].

Cada una de las etapas del ciclo de vida de la IA está a cargo de seres humanos responsables de definir sus características, objetivos, reglas de éxito, métricas y estructuras de análisis, así como de elegir los sets de datos para su entrenamiento, validación y testeo, lo que lleva ínsito el riesgo de que esas personas reflejen sus malas intenciones, así como sus propios valores, prejuicios y sesgos en el sistema[38].

Además, estos sistemas obtienen su conocimiento de las estructuras y dinámicas de la sociedad que produce los datos que los nutren, por lo que también pueden reproducir, reforzar y amplificar, con relativa facilidad, los patrones de exclusión, desigualdad y discriminación existentes[39].

El empleo cada vez mayor de algoritmos para automatizar la toma decisiones despierta también preocupación sobre los resultados discriminatorios que pueden provocar las decisiones o elecciones que hace la IA.

Se ha observado, por ejemplo, como los sets de datos utilizados para entrar modelos de IA usualmente excluyen a las personas adultas mayores o bien las representan estereotipadamente como sujetos sin habilidades e intereses diversos[40], con lo cual es muy probable que los sistemas entrenados a partir de esos datos no funcionen o funcionen peor con este grupo etario.

Asimismo, existe evidencia de la que herramienta de *Google Translate* puede exhibir sesgos de género que se derivan de los datos del mundo real con los que se entrena el sistema, esto es, de la forma en que nuestra sociedad escribe y habla sobre las mujeres[41].

También cómo el sistema COMPAS para predecir el riesgo de reincidencia de sujetos arrestados en los Estados Unidos, predijo valores de riesgo más elevados para personas de color y menos elevado para personas blancas en comparación con el riesgo real de cada sujeto[42].

En 2018, la agencia Reuters reportó que *Amazon* había desarrollado una herramienta inteligente para la selección de personal que, a partir de aprender de las preferencias pasadas en la materia, podía realizar un escaneo de los currículums recibidos tras una determinada convocatoria para

---

[37] Ampliar en Cathy O´NEIL, "La era de la fe ciega en los algoritmos debe terminar", abril de 2017, https://www.ted.com/talks/cathy_o_neil_the_era_of_blind_faith_in_big_data_must_end?language=es [accedido el 18/3/2022].

[38] Ver Cathy O'NEIL, "La era de la fe…". También: Andreas TSAMADOS, Nikita AGGARWAL, Josh COWLS, Jessica MORLEY, Huw ROBERTS, María Rosaria TADDEO y Luciano FLORIDI, "The ethics of algorithms…", p. 245.

[39] Cfr. Karen HAO, "This is how AI bias really happens – and why is so hard to fix", *MIT Technology Review*, 4 de febrero de 2019, en https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/ [accedido el 10/4/2022].

[40] WORLD HEALTH ORGANIZATION, "Ageism in artificial intelligence for health: WHO policy brief", 2022, p. 6, en https://www.analyticsinsight.net/ageism-in-artificial-intelligence-here-are-the-ways-to-prevent-,it/#:~:text=Ageism%20in%20AI%20is%20one%20new%20dimension%20to,AI%20and%20maximize%20AI%E2%80%99s%20benefits%20for%20older%20people [accedido el 15/5/2022].

[41] Marcelo PRATES, Pedro AVELAR, y Luis LAMBS, "Assessing gender bias in machine translation: a case study with Google Translate", *Neural Comput & Applic 32*, 27 de marzo de 2019, p. 48, en https://link.springer.com/article/10.1007/s00521-019-04144-6#citeas, accedido el [5/5/2022].

[42] Eirini NTOUTSI,Pavlos FAFALIOS,Ujwal GADIRAJU,Vasileios IOSIFIDIS,Wolfgang NEJDL,Maria-Esther VIDAL,Salvatore RUGGIERI,Franco TURINI,Symeon PAPADOPOULOS,Emmanouil KRASANAKIS,Ioannis KOMPATSIARIS,Katharina KINDER-KURLANDA,Claudia WAGNER,Fariba KARIMI,Miriam FERNANDEZ,Harith ALANI,Bettina BERENDT,Tina KRUEGEL,Christian HEINZE,Klaus BROELEMANN,Gjergji KASNECI,Thanassis TIROPANIS,Steffen STAAB, "Bias in data-driven artificial intelligence systems—An introductory survey", 3 de febrero de 2020, p. 2, en https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1356 [accedido el 12/5/2022].

identificar a los mejores candidatos comparando sus cualificaciones con las de candidatos contratados en el pasado.

El problema fue que el sistema rápidamente comenzó a puntuar peor aquellos currículums provenientes de candidatas mujeres y aún cuando *Amazon* intentó solucionar el problema, terminó por abandonar el uso del sistema al no poder asegurar que no estuviera sesgado contra las mujeres[43].

Por otra parte, también se ha observado como aplicaciones de reconocimiento de imágenes como *Amazon Rekognition* no etiquetan del mismo modo los elementos de una fotografía cuando aparecen hombres que cuando aparecen mujeres, pues reconocen un taladro como herramienta en una fotografía en la que hay un hombre y en cambio, cuando en la misma fotografía aparece una mujer, esa herramienta no es reconocida por el algoritmo[44].

3.3. La problemática de los *proxies*.

En muchas ocasiones los algoritmos no discriminan con base en alguna categoría o motivo prohibido, sino que lo hacen a partir de la consideración de atributos aparentemente neutros que, sin embargo, terminan por provocar una diferencia de trato perjudicial a miembros de determinadas clases o grupos protegidos.

Esto es lo que se ha dado en llamar la discriminación por *proxy*, que describe aquellas situaciones en las cuales el sistema utiliza una variable aparentemente neutra que termina por capturar el poder predictivo de una categoría prohibida de distinción[45].

En otras palabras: la IA puede provocar un trato desigual a partir de que adopta decisiones sobre la base de variables aparentemente neutrales que sirven de *proxies* para acceder a otras variables que han sido deliberadamente omitidas (v.gr. el género, la raza), dado que las primeras presentan una fuerte correlación con aquellas cuyo empleo se había descartado por no ser criterios de distinción legítimos, provocando el mismo efecto discriminatorio, pero sin que ello sea advertible a simple vista.

La discriminación por *proxies* puede ser el resultado de una mera coincidencia, de una discriminación estructural profundamente arraigada, o bien puede ser intencional. Esto último se denomina *masking* y ocurre cuando una variable proxy trivial o presuntamente neutral se emplea para enmascarar una intensión de discriminación consciente con base en una categoría protegida[46].

De esta manera, queda en claro que la simple omisión de ciertos datos personales durante el proceso de desarrollo de un sistema inteligente, como información relativa con el género o con la raza, no garantiza que se evite la discriminación. Esto es, si bien se reduce la posibilidad de

---

[43] Sarah MYERS WEST, *Discriminating Systems: Gender, Race, and Power in AI*, AI Now Institute, 2019, p. 8, https://ainowinstitute.org/discriminatingsystems.html [accedido el 12/5/2022].
[44] "Sexismo en la inteligencia artificial", https://www.youtube.com/watch?v=PkBdBDwVvJs.
[45] Anya PRINCE y Daniel SCHWARCZ, "Proxy Discrimination in the Age of Artificial Intelligence and Big Data", *Iowa Law Review* 105, 2020, p. 1263, en https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1695&context=faculty_articles [accedido ek 15/5/2022]. En el mismo sentido: Raphaële XENIDIS, "Tuning EU Equality Law to Algorithmic Discrimination: Three Pathways to Resilience", *Maastricht Journal of European and Comparative Law 27*, n.º 6, diciembre de 2020, p. 11, en https://journals.sagepub.com/doi/10.1177/1023263X20982173 [accedido el 12/5/2022].
[46] Janneke GERARDS y Raphaele XENIDIS, "Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law", *European Commission*, 10 de marzo de 2021, p. 44, en https://op.europa.eu/en/publication-detail/-/publication/082f1dbc-821d-11eb-9ac9-01aa75ed71a1 [accedido el 12/5/2022]. También ver Solon BAROCAS y Andrew SELBST, "Big Data's Disparate Impact", *California Law Review*, n.º 104, 30 de septiembre de 2016, p. 692, en https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899# [accedido el 12/5/2022].

incurrir en discriminación directa, todavía queda lugar para este tipo de discriminación indirecta y encubierta, que por ello se ha dado en llamar "el desafío de los proxies"[47].

Este desafío describe el hecho de que aun cuando los algoritmos pueden entrenarse para rechazar por irrelevantes criterios de distinción protegidos al encontrar patrones y correlaciones, igualmente pueden detectar aquellas variables *proxy* que están íntimamente relacionadas con la categoría protegida[48].

Lo complejo es que los *proxies* de atributos protegidos no son fáciles de predecir o detectar, en particular cuando los algoritmos acceden a sets de datos vinculados, en los que ciertos perfiles construidos a partir de características neutrales, como un código postal, pueden superponerse inadvertidamente con otros perfiles relacionados con el género, la orientación sexual, etcétera[49].

A modo de ejemplo, un estudio del año 2009 demostró que era posible predecir la orientación sexual de un usuario a partir de un análisis de sus amistades de Facebook, pues el porcentaje de amigos de un usuario que se autoidentifican como hombres homosexuales es una variable que presenta fuerte correlación la orientación sexual de ese usuario[50].

Del mismo modo, el lenguaje puede revelar ciertas características relevantes como la clase social —a partir de la selección de palabras para comunicarse, el dialecto o sociolecto—, el nivel educativo, la competencia como hablante no nativo e identidades o preferencias sociales particulares —a partir de la jerga—[51].

Y algunas características como el tiempo trabajado o la cantidad de ausencias de un candidato a un empleo en la última década, puede dar cuenta de característica de distinción prohibidas como el género y la condición de madre, pues sin dudas serán variables neutras pero fuertemente relacionadas con mujeres de 25 a 45 años que son madres de niños pequeños.

4. Sesgo y discriminación algorítmica. Aclaraciones conceptuales relevantes desde la perspectiva del Sistema Interamericano de Derechos Humanos.

4.1. ¿Qué es el sesgo algorítmico?

Es importante tener presente que la noción de *bias*, comúnmente empleada en la literatura en idioma inglés para abordar los riesgos de la IA para la igualdad y no discriminación, presenta algunas complejidades debido a los múltiples significados que tiene según el área de estudio en la que se la utilice —v.gr. ciencias de la computación, estadística, derecho—[52], así como también en razón de que su traducción al español, redunda en múltiples salidas como: sesgo, prejuicio, parcialidad, tendencia e inclinación, todos conceptos que, en nuestra lengua, a su vez, pueden tener alcance diverso.

---

[47] Janneke GERARDS y Raphaele XENIDIS, "Algorithmic discrimination in Europe: Challenges and opportunities…", p. 44/45.
[48] *Ídem*, p. 45.
[49] Bert HEINRICHS, "Discrimination in the age of artificial intelligence", *AI & Soc 37*, 4 de abril de 2021, p. 151, en https://link.springer.com/article/10.1007/s00146-021-01192-2#citeas [accedido el 12/5/2022].
[50] Frederik ZUIDERVEEN BORGESIUS, "Discrimination, artificial intelligence and algorithmic decision-making", *Directorate General of Democracy Council of Europe*, 2018, p.20/21, en https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73 [accedido el 25/5/2022].
[51] Laura WEIDINGER, John MELLOR, Maribeth RAUH, Conor GRIFFIN, Jonathan UESATO, Po-Sen HUANG, Myra CHENG, Mia GLAESE, Borja BALLE, Atoosa KASIRZADEH, Zac KENTON, Sasha BROWN, Will HAWKINS, Tom STEPLETON, Courtney BILES, Abeba BIRHANE, Julia HAAS, Laura RIMELL, Lisa Anne HENDRICKS, William ISAAC, Sean LEGASSICK, Geoffrey IRVING, Iason GABRIEL, "Ethical and social risks of harm from Language Models", 8 de diciembre de 2021, p. 18, en https://www.deepmind.com/publications/ethical-and-social-risks-of-harm-from-language-models [accedido el 12/5/2022].
[52] *Ídem*, p. 9.

Esta circunstancia conlleva la necesidad de formular algunas precisiones respecto del alcance del término a los fines de su empleo en el contexto de análisis señalado anteriormente.

Así, desde una óptica general, el concepto de "sesgo algorítmico" hace referencia a errores sistemáticos de cualquier tipo que se producen en el resultado de operaciones algorítmicas, lo que incluye aquellos de tipo estadístico, cognitivos, sociales, estructurales o de naturaleza institucional[53].

De hecho, desde una óptica estadística, el sesgo algorítmico se puede describir como aquella diferencia que se produce entre lo que el sistema predice y lo que sucede en la realidad o, en otras palabras, la que se produce entre el valor estimado o predicho y el verdadero valor[54].

En esta misma línea, en una investigación del año 2021, el Laboratorio de Inteligencia Artificial de la Universidad de Buenos Aires —IALAB— conceptualizó a los sesgos algorítmicos como aquellas decisiones erradas propias de un sistema informático[55].

Ahora bien, desde una perspectiva jurídica, en cambio, vinculada con el principio de igualdad y no discriminación, el "sesgo algorítmico" es una noción que se emplea para describir la inclinación o prejuicio de una decisión realizada por un sistema inteligente que afecta a una persona o un grupo determinado de una forma que se considera injusta o irrazonable[56].

En otras palabras, es un tipo de carga de valores en los sistemas inteligentes que da lugar a resultados injustos, sea porque coloca en desventaja a ciertos grupos o personas: (i) en función de características protegidas —sexo, raza, etnia, etc.—; (ii) debido a determinadas características no protegidas —código postal, auto que posee, largo de su apellido, etc.—; o (iii) de forma aleatoria[57].

4.2. El principio de igualdad y no discriminación en el Sistema Interamericano de Derechos Humanos.

La no discriminación, junto con la igualdad ante la ley y la igual protección de la ley en favor de todas las personas, "…son elementos constitutivos de un principio básico y general relacionado con la protección de los derechos humanos…"[58].

---

[53] Janneke GERARDS y Raphaele XENIDIS, "Algorithmic discrimination in Europe: Challenges and opportunities…", p. 47.

[54] Lindsey ANDERSEN, "Human Rights in the age of artificial intelligence", noviembre de 2018, p. 11, en https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf [accedido el 12/5/2022].

[55] Cfr. Cecilia DANESI, Laura DÍAZ DÁVILA, Mariana SÁNCHEZ CAPARRÓS y Pamela TOLOSA, "Sesgos algorítmicos de género", en Juan G. CORVALÁN y Nieves MACCHIAVELLI (Dir.) y Carolina CINALLI (Coord.), IALAB, 2021, en https://ialab.com.ar/wp-content/uploads/2022/03/Sesgos-algoritmicos-de-genero.pdf [accedido el 6/5/2022].

[56] Eirini NTOUTSI,Pavlos FAFALIOS,Ujwal GADIRAJU,Vasileios IOSIFIDIS,Wolfgang NEJDL,Maria-Esther VIDAL,Salvatore RUGGIERI,Franco TURINI,Symeon PAPADOPOULOS,Emmanouil KRASANAKIS,Ioannis KOMPATSIARIS,Katharina KINDER-KURLANDA,Claudia WAGNER,Fariba KARIMI,Miriam FERNANDEZ,Harith ALANI,Bettina BERENDT,Tina KRUEGEL,Christian HEINZE,Klaus BROELEMANN,Gjergji KASNECI,Thanassis TIROPANIS,Steffen STAAB, "Bias in data-driven artificial…", p. 3.

[57] Philip JANSEN y Philip BREY, "Ethical Analysis of AI…", p. 48.

[58] CORTE IDH, "Condición jurídica y derechos de los migrantes indocumentados", Opinión Consultiva OC-18/03, 17 de septiembre de 2003, párr. 83, en https://cdh.defensoria.org.ar/normativa/opinion-consultiva-oc-18-03-condicion-juridica-y-derechos-de-los-migrantes-indocumentados/ [accedido el 6/5/2022].

Así, el principio de igualdad y no discriminación aparece como un principio fundamental para la tutela de los derechos humanos, tanto en el derecho internacional como en el interno[59], que permea todo el ordenamiento jurídico[60].

Se trata de un principio que brinda "…una protección particularmente significativa, que incide en la garantía de todos los demás derechos y libertades consagrados en el derecho interno y el derecho internacional…"[61].

En el ámbito Latinoamericano, el artículo 1.1 de la Convención Americana sobre Derechos Humanos establece la obligación de los Estados de respetar y garantizar, sin discriminación alguna, los derechos y libertades allí reconocidos. De este modo, "…cualquiera sea el origen o la forma que asuma, todo tratamiento que pueda ser considerado discriminatorio respecto del ejercicio de cualquiera de los derechos garantizados en la Convención es *per se* incompatible con la misma…"[62].

El artículo 24 de la Convención, por su parte, "…prohíbe la discriminación de derecho o de hecho, no sólo en cuanto a los derechos consagrados en dicho tratado, sino en lo que respecta a todas las leyes que apruebe el Estado y a su aplicación…"[63].

Es importante tener presente que no toda diferencia de trato se considerará discriminatoria, "…sino sólo aquella que se base en criterios que no puedan ser racionalmente apreciados como objetivos y razonables…"[64].

La Corte Interamericana no ha dado una definición al término "discriminación" pero en sus precedentes ha invocado aquella dada por el Comité de Derechos Humanos del Pacto Internacional de Derechos Civiles y Políticos, que la ha descripto como: "…toda distinción, exclusión, restricción o preferencia que se basen en determinados motivos, como la raza, el color, el sexo, el idioma, la religión, la opinión política o de otra índole, el origen nacional o social, la propiedad, el nacimiento o cualquier otra condición social, y que tengan por objeto o por resultado anular o menoscabar el reconocimiento, goce o ejercicio, en condiciones de igualdad, de los derechos humanos y libertades fundamentales de todas las personas"[65].

Así, puede decirse que una diferencia de trato será discriminatoria cuando carezca de una justificación objetiva y razonable, es decir, cuando no persiga un fin legítimo y no exista una razonable proporción entre los medios utilizados y el fin perseguido[66].

Frente al trato diferente desfavorable, cuando el criterio diferenciador se corresponda "…con uno de aquellos protegidos por el artículo 1.1 de la Convención que aluden a: i) rasgos permanentes

---

[59] *Ídem*, párr. 88.

[60] CORTE IDH, "Caso Comunidad Indígena Xákmok Kásek Vs. Paraguay", sentencia del 24 de agosto de 2010, párr. 269, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_214_esp.pdf [accedido el 5/5/2022]. También Corte IDH, "Caso Yatama Vs. Nicaragua", sentencia del 23 de junio de 2005, párr. 184, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_127_esp.pdf [accedido el 5/5/2022].

[61] COMISIÓN IDH, Informe No. 04/01, "Caso 11.625, Fondo, María Eugenia Morales de Sierra, Guatemala", 19 de enero del 2001, párr. 163, en https://www.cidh.oas.org/annualrep/2000sp/capituloiii/fondo/Guatemala11.625.htm [accedido el 5/5/2022].

[62] CORTE IDH, "Caso Flor Freire Vs. Ecuador", sentencia del 31 de agosto de 2016, párr. 112, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_315_esp.pdf [accedido el 5/5/2022].

[63] CORTE IDH, "Caso Yatama Vs. Nicaragua", sentencia del 23 de junio de 2005, párr. 186, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_127_esp.pdf [accedido el 5/5/2022].

[64] CORTE IDH, "Caso I.V. Vs. Bolivia", sentencia del 30 de noviembre de 2016, párr. 240, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_329_esp.pdf [accedido el 5/6/2022].

[65] CORTE IDH, "Caso Atala Riffo y niñas Vs. Chile", sentencia del 24 de febrero de 2012, párr. 81, en https://corteidh.or.cr/docs/casos/articulos/seriec_239_esp.pdf [accedido el 4/6/2022].

[66] CORTE IDH, "Caso Norín Catrimán y otros (Dirigentes, miembros y activista del Pueblo Indígena Mapuche) Vs. Chile", sentencia del 29 de mayo de 2014, párr. 200, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_279_esp.pdf [accedido el 5/5/2022].

de las personas de los cuales éstas no pueden prescindir sin perder su identidad; ii) grupos tradicionalmente marginados, excluidos o subordinados, y iii) criterios irrelevantes para una distribución equitativa de bienes, derechos o cargas sociales…", la Corte entiende que se estará ante un indicio de que el Estado ha obrado con arbitrariedad[67].

Es dable recordar que, en punto a los criterios enunciados en el artículo 1.1 citado, que éstos no constituyen un listado taxativo o limitativo sino meramente enunciativo, por lo que se entiende que su redacción, al incluir la referencia a toda "otra condición social", permite incorporar otras categorías que no hubiesen sido explícitamente indicadas, pero que tengan una entidad asimilable[68].

En cualquier caso, al momento de interpretar aquel término "…deberá escogerse la alternativa hermenéutica más favorable a la tutela de los derechos de la persona humana, conforme a la aplicación del principio pro persona…"[69].

La discriminación puede, a su vez, ser de tipo directa o indirecta[70]. La primera ocurre cuando la diferencia de trato hacia una persona o grupo en análogas o similares situaciones se produce con base en una característica identificable del grupo protegida por las normas[71].

Cuando se trate de supuestos de discriminación directa, es decir, cuando la persona o grupo es tratado de forma menos favorable debido a pertenecer a algunas de las categorías protegidas del artículo 1.1. de la Convención, para probar la discriminación bastará con demostrar que la distinción legal emplea un factor prohibido o que no se cumplió con la acción positiva mandada por la norma[72].

Por su parte, la discriminación indirecta es aquella que ocurre cuando el impacto desproporcionado de normas, acciones, políticas u otras medidas, que son o parecen ser neutrales en su formulación o tienen un alcance general y no diferenciado, producen efectos negativos en una persona o grupo con características determinadas[73].

Ha destacado la Corte Interamericana de Derechos Humanos que el concepto de discriminación indirecta también ha sido reconocido, entre otros órganos, por su par europeo, que la descripto como aquella que ocurre "…cuando una política general o medida tiene un efecto desproporcionado perjudicial en un grupo particular (…) aún si no fue dirigida específicamente a ese grupo…"[74].

Frente a hipótesis de discriminaciones indirectas será necesario acreditar "…el efecto o resultado desproporcionadamente perjudicial que tiene ese criterio sobre un grupo o colectivo. En este

---

[67] CORTE IDH, "Opinión Consultiva OC-24/17", 24 de noviembre del 2017, párr. 66, en https://www.corteidh.or.cr/docs/opiniones/seriea_24_esp.pdf [accedido el 5/5/2022].
[68] *Ibidem.*
[69] *Ibidem.*
[70] CORTE IDH, "Condición jurídica y derechos de los migrantes indocumentados", Opinión Consultiva OC-18/03 del 17 de septiembre de 2003, párr. 103, en https://www.acnur.org/fileadmin/Documentos/BDL/2003/2351.pdf [accedido el 4/5/20220.]
[71] CORTE EDH, "Case Biao v. Denmark (Grand Chamber)", 24 de mayo del 2016, párr. 89, en https://hudoc.echr.coe.int/fre#{%22itemid%22:[%22001-163115%22]} [accedido el 5/5/2022].
[72] CORTE IDH, "Informe temático: Acceso a la justicia para las mujeres víctimas de violencia en las Américas", OEA/Ser.L/V/II, Documento 68, 20 de enero del 2007, párr. 91, en https://www.cidh.oas.org/pdf%20files/Informe%20Acceso%20a%20la%20Justicia%20Espanol%20020507.pdf, [accedido el 6/5/2022].
[73] CORTE IDH, "Caso Nadege Dorzema y otros Vs. República Dominicana", sentencia del 24 de octubre de 2012, párr. 235, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_251_esp.pdf [accedido el 3/5/2022]. También Corte IDH, "Caso Artavia Murillo y otros ("Fecundación in vitro") Vs. Costa Rica", sentencia de 28 noviembre de 2012, párr. 286, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_257_esp.pdf [accedido el 5/5/2022].
[74] CORTE IDH, "Caso Nadege Dorzema y otros Vs. República Dominicana", sentencia del 24 de octubre de 2012, párr. 235.

caso, acreditar la existencia de discriminación supone aportar datos empíricos que demuestren que el supuesto sesgo "invisible" o "neutral" en la adopción de decisiones tiene un efecto dispar sobre algún grupo o sobre grupos determinados…”[75].

4.3. Recapitulación.

Dijimos que el sesgo, desde una mirada relacionada con el principio de igualdad y no discriminación, es un tipo de carga de valores en los sistemas inteligentes que da lugar a resultados injustos, sea porque coloca en desventaja a ciertos grupos o personas: (i) en función de características protegidas —sexo, raza, etnia, etc.—; (ii) debido a determinadas características no protegidas —código postal, auto que posee, largo de su apellido, etc.—; o (iii) de forma aleatoria[76].

Desde esta perspectiva, se trata de un concepto que puede emplearse como sinónimo del de discriminación algorítmica en sentido amplio, en tanto esta última, tomando como base la conceptualización de la discriminación que realiza la Corte Interamericana de Derechos Humanos[77], puede definirse como aquella que se produce cuando los sistemas inteligentes provocan un impacto o trato desfavorable respecto de ciertas personas o grupos, en base a criterios o características que no pueden ser racionalmente apreciados como objetivos y razonables.

El sesgo algorítmico, en tanto concepto amplio empleado desde la perspectiva señalada, resulta comprensivo tanto de la noción de discriminación directa como de la de discriminación indirecta.

La primera, cabe recordar, es aquella que tiene lugar cuando una persona es tratada de forma menos favorable debido a pertenecer a alguna de las categorías de discriminación vedada contenidas en el artículo 1.1 de la Convención Americana[78].

Así, la discriminación directa recoge situaciones en las que una decisión es adoptada tomando en cuenta alguna característica protegida en perjuicio de las personas o grupos relacionados con ese motivo protegido[79]. Para probar la discriminación directa habrá que demostrar que la diferencia de trato cuestionada emplea un factor prohibido o que no se cumplió con la acción positiva mandada por la norma[80].

La discriminación indirecta, por su parte, es aquella que sucede cuando normas, acciones, políticas u otras medidas, que son o parecen ser neutrales en su formulación o tienen un alcance general y no diferenciado, producen impactos negativos desproporcionados en una persona o grupo con características determinadas[81].

Es decir, cuando una disposición, criterio o práctica aparentemente neutro coloca a los miembros de una categoría protegida en una situación de desventaja particular en comparación con otras

---

[75] CORTE IDH, “Informe temático: Acceso a la justicia para las mujeres…”.
[76] Philip JANSEN y Philip BREY, “Ethical Analysis of AI…”, p. 48.
[77] CORTE IDH, “Caso I.V. Vs. Bolivia”, sentencia del 30 de noviembre de 2016, párr. 240, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_329_esp.pdf [accedido el 5/6/2022].
[78] CORTE IDH, “Informe temático: Acceso a la justicia para las mujeres víctimas…”.
[79] Janneke GERARDS y Raphaele XENIDIS, “Algorithmic discrimination in Europe: Challenges and opportunities…”, p. 68.
[80] CORTE IDH, “Informe temático: Acceso a la justicia para las mujeres víctimas…”.
[81] CORTE IDH, “Caso Nadege Dorzema y otros Vs. República Dominicana”, sentencia del 24 de octubre de 2012, párr. 235, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_251_esp.pdf [accedido el 3/5/2022]. También CORTE IDH, “Caso Artavia Murillo y otros ("Fecundación in vitro") Vs. Costa Rica”, sentencia de 28 noviembre de 2012, párr. 286, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_257_esp.pdf [accedido el 5/5/2022].

personas, y esa disposición, criterio o práctica no persigue un fin legítimo y no existe una razonable proporción entre los medios utilizados y el fin perseguido[82].

En el contexto de la IA el concepto de discriminación indirecta es el que mejor se adapta a la lógica de la discriminación algorítmica, pues en lugar de centrarse en cómo se trata a los individuos o al grupo debido a sus características, se enfoca en los efectos que una medida o decisión del sistema provoca en los grupos protegidos[83].

En tanto la discriminación indirecta pone el foco en los efectos discriminatorios de los sistemas más en que en su operatoria, resulta un concepto más útil para sortear algunas dificultades relacionadas con el acceso al contenido de los algoritmos y la comprensión de cómo operan por parte de no expertos, así como las relativas a la responsabilidad que puede caber a los diferentes actores que intervienen durante el ciclo de vida de la IA[84].

Además, es un concepto que permite capturar adecuadamente los supuestos de discriminación por *proxies* —tratado en el punto 3.3—, referido a situaciones en las que las decisiones del sistema provocan un impacto desigual con base en variables aparentemente neutras que están correlacionadas con motivos de distinción vedados[85].

5. Conclusiones.

La toma de decisiones automatizadas se observa cada vez más en dominios sociales sensibles como la justicia, la medicina, las finanzas y el sector público. Pero de la mano de ello y a raíz de los sesgos que afectan los sistemas inteligentes que sirven de vehículo para esa automatización, surgen nuevas formas discriminación que afectan a las personas.

En ese marco, es irresponsable —y hasta peligroso— que la noción de “decisión basada en datos” se presente a la sociedad asociada a la idea de no discriminatorio, no sesgado u objetivo[86].

Básicamente porque por el importante rol que están llamadas a cumplir las personas que intervienen en el ciclo de vida de los sistemas inteligentes, siempre existe la posibilidad real de que sus propios prejuicios y sesgos se trasladen a los sistemas[87]. Además, ello puede suceder también como consecuencia de los datos con los que se los entrena.

Los sistemas inteligentes son herramientas diseñadas para predecir a partir de efectuar correlaciones de entre los conjuntos de datos de entrenamiento, no para efectuar inferencias causales. Esta limitación implica que las elecciones pasadas, reflejadas en esos datos, van a tener un efecto importante en la predicción, pues son determinantes para los resultados del sistema[88].

La IA depende de los datos generados por los seres humanos, por lo que, sin los cuidados e intervenciones adecuados, cualquier prejuicio que exista en las personas o las comunidades que los producen ingresará en los sistemas empleándolos como vehículo. Como resultado, éstos

---

[82] CORTE IDH, “Caso Norín Catrimán y otros (Dirigentes, miembros y activista del Pueblo Indígena Mapuche) Vs. Chile”, sentencia del 29 de mayo de 2014, párr. 200, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_279_esp.pdf [accedido el 5/5/2022].
[83] Janneke GERARDS y Raphaele XENIDIS, “Algorithmic discrimination in Europe: Challenges and opportunities…”, p. 70.
[84] *Idem*, p. 70.
[85] *Ibidem*.
[86] Jon KLEINBERG, Jens LUDWIG, Sendhil MULLAINATHAN, Cass R. SUNSTEIN, “Discrimination in the Age of Algorithms”, 5 de febrero de 2019, p. 18, en https://arxiv.org/ftp/arxiv/papers/1902/1902.03731.pdf [accedido el 15/5/2022].
[87] *Ibidem*.
[88] *Idem*, p.21.

últimos habrán de reproducir o amplificar las desigualdades reflejadas en los conjuntos de datos con los que fueron entrenados[89].

De allí lo complejo, pues al igual que los prejuicios del mundo real, el sesgo algorítmico no se encuentra en un único lugar como un error que se puede localizar fácilmente y corregir[90].

Desafíos adicionales se presentan con la cuestión de los *proxies*, pues los datos codifican las características de las personas en forma de valores o atributos que pueden estar asociados o servir para identificar motivos de distinción vedados. De esta manera, eliminar o ignorar atributos sensibles no impide que los modelos de aprendizaje sean igualmente sesgados, ya que otras características poco redundantes y aparentemente neutrales pueden ser empleadas como *proxies* hacia ellos[91].

Ahora bien, los seres humanos siguen siendo los responsables de preparar y etiquetar los datos que alimentarán los sistemas; de proveer la retroalimentación durante el proceso de aprendizaje iterativo; de hacer la selección de los atributos relevantes y definir los objetivos del sistema, así como de hacer el control de calidad y la validación de sus resultados. Por esta razón, el rol de las personas durante el ciclo de vida de la IA se torna crucial para abordar los riesgos que ésta representa para el principio de igualdad y no discriminación.

Ello puesto que, si bien se han propuesto y diseñado diversas herramientas de tipo técnico para abordar la problemática del sesgo de los sistemas inteligentes, lo cierto es que todavía no existe un entendimiento suficiente acerca de cómo resguardar sin error valores humanos críticos como la seguridad o la equidad cuando se los diseña, desarrolla y despliega[92].

Por esta razón, de un tiempo a esta parte se ha advertido sobre la necesidad de hacer hincapié y promover la intervención y supervisión humana constante de los sistemas mediante equipos interdisciplinarios y socialmente conscientes, así como integrar al campo de investigación de la IA el enorme historial de investigaciones en materia de igualdad y no discriminación, originadas en las ciencias sociales y las humanidades[93].

Es que la problemática de la discriminación no es nueva, es tan antigua como la civilización humana, pues históricamente ha estado en la naturaleza de los miembros de mayorías dominantes el ser ajenos a las experiencias de otros grupos sociales[94].

---

[89] Eirini NTOUTSI,Pavlos FAFALIOS,Ujwal GADIRAJU,Vasileios IOSIFIDIS,Wolfgang NEJDL,Maria-Esther VIDAL,Salvatore RUGGIERI,Franco TURINI,Symeon PAPADOPOULOS,Emmanouil KRASANAKIS,Ioannis KOMPATSIARIS,Katharina KINDER-KURLANDA,Claudia WAGNER,Fariba KARIMI,Miriam FERNANDEZ,Harith ALANI,Bettina BERENDT,Tina KRUEGEL,Christian HEINZE,Klaus BROELEMANN,Gjergji KASNECI,Thanassis TIROPANIS,Steffen STAAB, "Bias in data-driven artificial intelligence systems—An introductory survey", 3 de febrero de 2020, p. 3, https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1356.

[90] Myers West, *Discriminating Systems: Gender, Race, and Power in AI*, p. 19.

[91] Eirini NTOUTSI,Pavlos FAFALIOS,Ujwal GADIRAJU,Vasileios IOSIFIDIS,Wolfgang NEJDL,Maria-Esther VIDAL,Salvatore RUGGIERI,Franco TURINI,Symeon PAPADOPOULOS,Emmanouil KRASANAKIS,Ioannis KOMPATSIARIS,Katharina KINDER-KURLANDA,Claudia WAGNER,Fariba KARIMI,Miriam FERNANDEZ,Harith ALANI,Bettina BERENDT,Tina KRUEGEL,Christian HEINZE,Klaus BROELEMANN,Gjergji KASNECI,Thanassis TIROPANIS,Steffen STAAB, "Bias in data-driven artificial…", p. 4.

[92] Roel DOBBE, Sarah DEAN, Thomas GILBERT, Nitin KOHLI, "A Broader View on Bias in Automated Decision-Making: Reflecting on Epistemology and Dynamics", presentado en el *Workshop on Fairness, Accountability and Transparency in Machine Learning* during ICML 2018, Stockholm, Sweden, 6 de julio de 2018), p. 42, en https://arxiv.org/abs/1807.00553#:~:text=version%2C%20v2)%5D-,A%20Broader%20View%20on%20Bias%20in%20Automated%20Decision%2DMaking,Reflecting%20on%20Epistemology%20and%20Dynamics&text=Machine%20learning%20(ML)%20is%20increasingly,of%20automated%20decision%2Dmaking%20systems [accedido el 12/5/2022].

[93] *Idem,* p. 42.

[94] Eirini NTOUTSI,Pavlos FAFALIOS,Ujwal GADIRAJU,Vasileios IOSIFIDIS,Wolfgang NEJDL,Maria-Esther VIDAL,Salvatore RUGGIERI,Franco TURINI,Symeon PAPADOPOULOS,Emmanouil

Lo que sí es nuevo, es que con la introducción de la IA se produce el riesgo, no sólo de que se magnifiquen y amplifiquen los prejuicios o diferencias de trato perjudiciales ya existentes, sino de que surjan nuevas clasificaciones con impacto igualmente negativo a partir de criterios que a primera vista parecen ser neutrales por lo que podrían ser ignorados hasta tanto provoquen un daño en algún sector de la sociedad que, lamentablemente, siempre suele coincidir con que es un grupo o clase vulnerable.

De allí que gran parte de la investigación en materia de IA se haya enfocado en la problemática del sesgo algorítmico y su impacto en el derecho a la igualdad y no discriminación de las personas.

No obstante, se requiere mayor investigación interdisciplinaria para definir qué es lo que se necesita para lograr un equilibrio entre los derechos y libertades fundamentales de los ciudadanos a partir de la mitigación de los sesgos algorítmicos, a la par de considerar cuáles son los desafíos técnicos y los requerimientos económicos para lograr ese fin[95].

Mientras tanto, es muy posible que, frente a la impugnación de decisiones automatizadas, la falta de transparencia acerca de cómo funcionan los sistemas inteligentes lleve a los tribunales a considerar que un sistema es *prima facie* discriminatorio, y coloque a quienes los cuestionan en mejor posición para salir exitoso de aquel cuestionamiento[96].

En dicho marco, las normas antidiscriminación contenidas en la Convención Americana sobre Derechos Humanos lucen, *prima facie*, aptas para proteger a las personas de la discriminación impulsada por la IA.

Dicha Convención establece la obligación de los Estados Partes de "…respetar los derechos y libertades reconocidos en ella y a garantizar su libre y pleno ejercicio a toda persona que esté sujeta a su jurisdicción, sin discriminación alguna por motivos de raza; color, sexo, idioma, religión, opiniones políticas o de cualquier otra índole, origen nacional o social, posición económica, nacimiento o cualquier otra condición social…"[97].

Así, este precepto resulta comprensivo, tanto de los supuestos de discriminación directa como indirecta, por lo que se visualiza adecuado para tutelar el derecho de las personas frente a situaciones de discriminación algorítmica.

Sobre esta última noción, es importante tener en cuenta que cuando se aborda la problemática en el trato, es frecuente que se haga referencia a la noción de "sesgo algorítmico".

Este concepto se presenta adecuado en la medida en que, siempre desde una óptica relacionada con el principio de igualdad y no discriminación, es una noción que refiere a aquella carga de valores en los sistemas inteligentes que da lugar a resultados injustos, sea porque coloca en desventaja a ciertos grupos o personas: (i) en función de características protegidas —sexo, raza, etnia, etc.—; (ii) debido a determinadas características no protegidas —código postal, auto que posee, largo de su apellido, etc.—; o (iii) de forma aleatoria[98].

Por ello, desde la perspectiva restringida señalada, puede emplearse como sinónimo del de discriminación algorítmica, en tanto esta última, partiendo de la conceptualización de la noción

---

KRASANAKIS,Ioannis KOMPATSIARIS,Katharina KINDER-KURLANDA,Claudia WAGNER,Fariba KARIMI,Miriam FERNANDEZ,Harith ALANI,Bettina BERENDT,Tina KRUEGEL,Christian HEINZE,Klaus BROELEMANN,Gjergji KASNECI,Thanassis TIROPANIS,Steffen STAAB, "Bias in data-driven artificial…", p. 2.

[95] Robin ALLEN QC, *Artificial Intelligence, machine learning, algorithms and discrimination law: the new frontier*, 31 de enero de 2020, p. 10, en https://482pe539799u3ynseg2hl1r3-wpengine.netdna-ssl.com/wp-content/uploads/2020/02/Discrimination-Law-in-2020.FINAL_-1.pdf [accedido el 15/5/2022].

[96] *Idem*, p. 18.

[97] Cfr. Art. 1.1 CADH.

[98] Philip JANSEN y Philip BREY, "Ethical Analysis of AI…", p. 48.

de discriminación que realiza la Corte Interamericana de Derechos Humanos[99], se puede describir como aquella que se produce cuando los sistemas inteligentes provocan un impacto o trato desfavorable respecto de ciertas personas o grupos, en base a criterios o características que no pueden ser racionalmente apreciados como objetivos y razonables.

Ahora bien, en el contexto de la IA el concepto de discriminación indirecta es el que mejor se adapta a la lógica de la discriminación algorítmica, puesto que se enfoca en los efectos que una medida o decisión del sistema provoca en grupos alcanzados por categorías protegidas[100].

Es decir, es una noción que se enfoca en los efectos discriminatorios de los sistemas más que en su operatoria. De allí su notable utilidad para sortear algunas dificultades relacionadas con el acceso al proceso ejecutado por los algoritmos y la comprensión de cómo operan por parte de no expertos, así como también las relativas a la responsabilidad que puede caber a los diferentes actores que intervienen a lo largo del ciclo de vida de la IA[101].

Por lo demás, no quedan dudas de que es el que mejor captura los supuestos de discriminación por *proxies* —tratado en el punto 3.3—, en los que los impactos desiguales se dan a partir de las correlaciones que realiza el sistema entre variables que a primera vista lucen neutras pero que sirven para acceder a otras variables protegidas, incluso si estas últimas fueron deliberadamente omitidas.

No obstante, se ha señalado que aún en la hipótesis de discriminación indirecta, al ser necesario demostrar que una práctica, norma o decisión aparentemente neutra afecta desproporcionadamente a un grupo protegido y, por ende, es en principio discriminatoria, este extremo puede dificultar el acceso a la justicia de las personas, en tanto les demanda un esfuerzo adicional de obtener evidencia estadística para mostrar el efecto de esa desproporción[102].

Es por ello que más allá de que la problemática de la no discriminación se aborde desde un enfoque de derechos individuales basado en el hecho de que sean los particulares los que impugnen la decisión automatizada que discrimina ante los tribunales, frente a las dificultades probatorias y de acceso a la justicia que este enfoque puede representar, se ha destacado que es importante pensar también en uno de supervisión, que coloque en cabeza de instituciones públicas el deber de tomar acciones dirigidas a reforzar la igualdad[103].

Finalmente, y sin perjuicio de lo expuesto, en vista a lo sucedido con iniciativas como el *chatbot Tay* de Microsoft o el buscador de personal de Amazon, no quedan dudas de que más allá de la existencia de alternativas ciertas de impugnar las decisiones automatizadas que discriminan, lo cierto es que determinados sistemas que se muestran seriamente sesgados no deberían tener la chance de ser arreglados o mejorados, sino que directamente tendrían que dejar de emplearse por completo[104].

---

[99] CORTE IDH, "Caso I.V. Vs. Bolivia", sentencia del 30 de noviembre de 2016, párr. 240, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_329_esp.pdf [accedido el 5/6/2022].
[100] Janneke GERARDS y Raphaele XENIDIS, "Algorithmic discrimination in Europe: Challenges and opportunities…", p. 70.
[101] *Ibidem*.
[102] Zuiderveen BORGESIUS, *Discrimination, artificial intelligence and algorithmic decision-making*, Directorate General of Democracy, Council of Europe, p. 34, en https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73 [accedido el 12/5/2022].
[103] Raphaële XENIDIS y Linda SENDEN, "EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination", 30 de septiembre de 2019, p. 23, en https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3529524 [accedido el 12/5/2022].
[104] Myers West, *Discriminating Systems: Gender, Race, and Power in AI,* p. 19.

Bibliografía:

ALLEN QC Robin, *Artificial Intelligence, machine learning, algorithms and discrimination law: the new frontier*, 31 de enero de 2020, en https://482pe539799u3ynseg2hl1r3-wpengine.netdna-ssl.com/wp-content/uploads/2020/02/Discrimination-Law-in-2020.FINAL_-1.pdf

ANDERSEN Lindsey, "Human Rights in the age of artificial intelligence", noviembre de 2018, en https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf

ASAMBLEA GENERAL DE LAS NACIONES UNIDAS, "Resolución 73/348", en http://undocs.org/es/A/73/348.

BAROCAS Solon y SELBST Andrew, "Big Data's Disparate Impact", *California Law Review*, n.º 104, 30 de septiembre de 2016, p. 671-732, en https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899#

BORGESIUS Zuiderveen, *Discrimination, artificial intelligence and algorithmic decision-making*, Directorate General of Democracy, Council of Europe, en https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73

BBC MUNDO, "Google pide perdón por confundir a una pareja negra con gorilas", 15 de julio de 2015, https://www.bbc.com/mundo/noticias/2015/07/150702_tecnologia_google_perdon_confundir_afroamericanos_gorilas_lv

BBC MUNDO, "5 claves para entender el escándalo de Cambridge Analytica que hizo que Facebook perdiera U$S 37.000 millones en un día", 20 de marzo de 2018, en https://www.bbc.com/mundo/noticias-43472797.

BBC MUNDO, "Tay, la robot racista y xenófoba de Microsoft", 25 de marzo de 2016, en https://www.bbc.com/mundo/noticias/2016/03/160325_tecnologia_microsoft_tay_bot_adolescente_inteligencia_artificial_racista_xenofoba_lb.

BBC MUNDO, "Qué es la 'teoría del empujón' que le hizo ganar al estadounidense Richard H. Thaler el Nobel de Economía", 9 de octubre de 2017, en https://www.bbc.com/mundo/noticias-41551856.

BOSTROM Nick, "The Ethics of Artificial Intelligence", *Draft for Cambridge Handbook of Artificial Intelligence*, Londres, 2011, en https://www.nickbostrom.com/ethics/artificial-intelligence.pdf.

COMISIÓN EUROPEA, *Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza*, Bruselas, 2020, en https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificialintelligence-feb2020_es.pdf.

COMISIÓN IDH, Informe No. 04/01, "Caso 11.625, Fondo, María Eugenia Morales de Sierra, Guatemala", 19 de enero del 2001, en https://www.cidh.oas.org/annualrep/2000sp/capituloiii/fondo/Guatemala11.625.htm.

COMISIÓN MUNDIAL DE ÉTICA DEL CONOCIMIENTO CIENTÍFICO Y LA TECNOLOGÍA, "Preliminary Study on the Ethics of Artificial Intelligence", Paris, UNESCO, 26 de febrero de 2019, en https://unesdoc.unesco.org/ark:/48223/pf0000367823.

CORTE EDH, "Case Biao v. Denmark (Grand Chamber)", 24 de mayo del 2016, en https://hudoc.echr.coe.int/fre#{%22itemid%22:[%22001-163115%22]}

Corte IDH, "Caso Artavia Murillo y otros ("Fecundación in vitro") Vs. Costa Rica", sentencia de 28 noviembre de 2012, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_257_esp.pdf.

CORTE IDH, "Caso Atala Riffo y niñas Vs. Chile", sentencia del 24 de febrero de 2012, en https://corteidh.or.cr/docs/casos/articulos/seriec_239_esp.pdf

CORTE IDH, "Caso I.V. Vs. Bolivia", sentencia del 30 de noviembre de 2016, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_329_esp.pdf.

CORTE IDH, "Caso Comunidad Indígena Xákmok Kásek Vs. Paraguay", sentencia del 24 de agosto de 2010, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_214_esp.pdf.

CORTE IDH, "Caso Flor Freire Vs. Ecuador", sentencia del 31 de agosto de 2016, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_315_esp.pdf.

CORTE IDH, "Caso Nadege Dorzema y otros Vs. República Dominicana", sentencia del 24 de octubre de 2012, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_251_esp.pdf.

CORTE IDH, "Caso Norín Catrimán y otros (Dirigentes, miembros y activista del Pueblo Indígena Mapuche) Vs. Chile", sentencia del 29 de mayo de 2014, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_279_esp.pdf.

CORTE IDH, "Condición jurídica y derechos de los migrantes indocumentados", Opinión Consultiva OC-18/03 del 17 de septiembre de 2003, en https://www.acnur.org/fileadmin/Documentos/BDL/2003/2351.pdf.

CORTE IDH, “Informe temático: Acceso a la justicia para las mujeres víctimas de violencia en las Américas”, OEA/Ser.L/V/II, Documento 68, 20 de enero del 2007, en https://www.cidh.oas.org/pdf%20files/Informe%20Acceso%20a%20la%20Justicia%20Espanol%20020507.pdf.
CORTE IDH, “Opinión Consultiva OC-24/17”, 24 de noviembre del 2017, en https://www.corteidh.or.cr/docs/opiniones/seriea_24_esp.pdf
Corte IDH, “Caso Yatama Vs. Nicaragua”, sentencia del 23 de junio de 2005, en https://www.corteidh.or.cr/docs/casos/articulos/seriec_127_esp.pdf.
CORTE IDH, “Condición jurídica y derechos de los migrantes indocumentados”, Opinión Consultiva OC-18/03, 17 de septiembre de 2003, en https://cdh.defensoria.org.ar/normativa/opinion-consultiva-oc-18-03-condicion-juridica-y-derechos-de-los-migrantes-indocumentados/.
CORVALÁN Juan G., *Perfiles Digitales Humanos*, La Ley, Buenos Aires, 2020
COUNCIL OF EUROPE, *A legal framework for AI systems. Feasibility study for a legal framework for the development, design and application of artificial intelligence, based on Council of Europe´s standards on human rights, democracy and the rule of law*, adoptado por la CAHAI en su 3ra reunion plenaria el 17 de diciembre de 2020, en https://media-exp1.licdn.com/dms/document/C4D1FAQEe4eGCLGf6CA/feedshare-document-pdf-analyzed/0/1649063227482?e=2147483647&v=beta&t=ODWWaRAOoF3tufI-aNt1wCP57PWKqf8mk8FkJ6alA58.
DANESI Cecilia, DÍAZ DÁVILA Laura, SÁNCHEZ CAPARRÓS Mariana y TOLOSA Pamela, “Sesgos algorítmicos de género”, en CORVALÁN Juan G. y MACCHIAVELLI Nieves (Dir.) y CINALLI Carolina (Coord.), IALAB, 2021, en https://ialab.com.ar/wp-content/uploads/2022/03/Sesgos-algoritmicos-de-genero.pdf.
DASTIN Jeffrey, “Amazon scraps secret AI recruiting tool that showed bias against women”, 10 de octubre de 2018, *Reuters*, disponible en www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-airecruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.
DOBBE Roel, DEAN Sarah, GILBERT Thomas y KOHLI Nitin, “A Broader View on Bias in Automated Decision-Making: Reflecting on Epistemology and Dynamics”, presentado en el *Workshop on Fairness, Accountability and Transparency in Machine Learning* during ICML 2018, Stockholm, Sweden, 6 de julio de 2018, en https://arxiv.org/abs/1807.00553#:~:text=version%2C%20v2)%5D-,A%20Broader%20View%20on%20Bias%20in%20Automated%20Decision%2DMaking,Reflecting%20on%20Epistemology%20and%20Dynamics&text=Machine%20learning%20(ML)%20is%20increasingly,of%20automated%20decision%2Dmaking%20systems
GERARDS Janneke y XENIDIS Raphaele, “Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law”, *European Commission*, 10 de marzo de 2021, en https://op.europa.eu/en/publication-detail/-/publication/082f1dbc-821d-11eb-9ac9-01aa75ed71a1
GRUPO DE EXPERTOS DE ALTO NIVEL SOBRE IA, “Directrices éticas para una IA fiable”, abril de 2019, en https://urjes.com/pruebas/wp-content/uploads/2021/05/Grupo-independiente-de-expertos-...-Directrices-eticas-para-una-inteligencia-artificial-IA-fiable-ES.pdf.
HAO Karen, “This is how AI bias really happens – and why is so hard to fix”, *MIT Technology Review*, 4 de febrero de 2019, en https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/.
HEINRICHS Bert, “Discrimination in the age of artificial intelligence”, *AI & Soc 37*, 4 de abril de 2021, p. 143-154, en https://link.springer.com/article/10.1007/s00146-021-01192-2#citeas
HEVIA MARTÍNEZ German, “La sociedad como artefacto. Sistemas sociotécnicos, sociotecnologías y sociotécnicas”, *Revista Iberoamericana de Ciencia Tecnología y Sociedad* 40, febrero de 2019, en https://www.researchgate.net/publication/332863180_La_sociedad_como_artefacto_Sistemas_sociotecnicos_sociotecnologias_y_sociotecnicas
HILL Robin K., “What an Algorithm Is”, *Philos. Technol*, 11 de mayo de 2015, en https://link.springer.com/article/10.1007/s13347-014-0184-5.
IENCA Marcelo y VAYENA Effy, “AI Ethics Guidelines: European and Global Perspectives”, en CAHAI Secretariat (Comp.), *Towards Regulation of AI Systems*, diciembre de 2020, p. 38-60, en https://www.coe.int/en/web/artificial-intelligence/-/-toward-regulation-of-ai-systems-.

INFOBAE, “Facebook lamenta etiqueta en video de hombres de raza negra”, 4 de septiembre de 2021, https://www.infobae.com/america/agencias/2021/09/05/facebook-lamenta-etiqueta-en-video-de-hombres-de-raza-negra/
JANSEN Philip y BREY Philip, “Ethical Analysis of AI and Robotics Technologies”, *SIENNA Project*, 31 de agosto de 2019, en https://www.sienna-project.eu/digitalAssets/801/c_801912-l_1-k_d4.4_ethical-analysis--ai-and-r--with-acknowledgements.pdf.
KLEINBERG Jon, LUDWIG Jens, MULLAINATHAN Sendhil, SUNSTEIN Cass R., “Discrimination in the Age of Algorithms”, 5 de febrero de 2019, p. 1-45, en https://arxiv.org/ftp/arxiv/papers/1902/1902.03731.pdf
LESLIE David, “Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector”, *The Alan Turing Institute*, 2019, en https://www.turing.ac.uk/research/publications/understanding-artificial-intelligence-ethics-and-safety [accedido el 15/3/2022].
LÓPEZ ONETO Marcos, “Fundamentos antropológicos, éticos, filosóficos, históricos, sociológicos y jurídicos para la constitución universal de un derecho de la inteligencia artificial (DIA)”, p. 69-158, en Juan G. CORVALÁN (Dir.), *Tratado de Inteligencia Artificial y Derecho*, La Ley, Buenos Aires, 2021, t. 1.
MYERS WEST Sarah, *Discriminating Systems: Gender, Race, and Power in AI*, AI Now Institute, 2019, https://ainowinstitute.org/discriminatingsystems.html.
METZ Rachel, “La IA racista y neonazi de Microsoft fue culpa de Microsoft”, *MIT Technology Review*, marzo 2016, en https://www.technologyreview.es/s/5720/la-ia-racista-y-neonazi-de-microsoft-fue-culpa-de-microsoft.
NTOUTSI Eirini, FAFALIOS Pavlos, GADIRAJU Ujwal, IOSIFIDIS Vasileios, NEJDL Wolfgang, VIDAL Maria-Esther, RUGGIERI Salvatore, TURINI Franco, PAPADOPOULOS Symeon, KRASANAKIS Emmanouil, KOMPATSIARIS Ioannis, KINDER-KURLANDA Katharina, WAGNER Claudia, KARIMI Fariba, FERNANDEZ Miriam, ALANI Harith, BERENDT Bettina, KRUEGEL Tina, HEINZE Christian, BROELEMANN Klaus, KASNECI Gjergji, TIROPANIS Thanassis, STAAB Steffen, “Bias in data-driven artificial intelligence systems—An introductory survey”, 3 de febrero de 2020, en https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1356.
O´NEIL Cathy, “La era de la fe ciega en los algoritmos debe terminar”, abril de 2017, https://www.ted.com/talks/cathy_o_neil_the_era_of_blind_faith_in_big_data_must_end?language=es.
PALERMO Joseph, “Deepfakes: Why you can´t believe everything you see and hear”, 12 de febrero de 2020, e n https://www.youtube.com/watch?v=JXBBalHl_cI.
PRATES Marcelo, AVELAR Pedro y LAMBS Luis, “Assessing gender bias in machine translation: a case study with Google Translate”, *Neural Comput & Applic 32*, 27 de marzo de 2019, p. 6363–6381, en https://link.springer.com/article/10.1007/s00521-019-04144-6#citeas.
PRINCE Anya y SCHWARCZ Daniel, “Proxy Discrimination in the Age of Artificial Intelligence and Big Data”, *Iowa Law Review* 105, 2020, p. 1257-1318, en https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1695&context=faculty_articles.
RETINA Tendencias, “¿Qué es y en qué consiste el sistema de crédito social chino?”, 5 de febrero de 2021, en https://www.youtube.com/watch?v=lun7V7TPmmM.
RIPANI Laura, KUGLER Adriana, SOLER Nicolás, KUGLER Maurice y RODRIGO Ramiro, *El futuro del trabajo en América Latina y el Caribe: ¿Cuál es el impacto de la automatización en el empleo y los salarios?*, Banco Interamericano de Desarrollo, diciembre de 2020, en https://publications.iadb.org/es/el-futuro-del-trabajo-en-america-latina-y-el-caribe-cual-es-el-impacto-de-la-automatizacion-en-el.
SALAS Javier, “El algoritmo que discrimina a los pacientes negros sin conocer su raza”, *El País*, 24 de octubre de 2019, en https://elpais.com/elpais/2019/10/24/ciencia/1571909798_596622.html
SAMOILI Sofia, LÓPEZ COBO Montserrat, DELIPETREV Blajob, MARTÍNEZ-PLUMED Fernando, GÓMEZ Emilia y Giuditta DE PRATO Giuditta, *AI Watch. Defining Artificial Intelligence 2.0: Towards an operational definition and taxonomy for the AI landscape*, Publications Office of the European Union, Luxembourg, 2021, en https://publications.jrc.ec.europa.eu/repository/handle/JRC118163.
SENRA Ricardo, “El dilema de las redes sociales’ de Netflix: 5 secretos de los dueños de las redes para engancharnos y manipularnos, según el documental”, *BBC News Brasil*, 3 de octubre de 2020, en https://www.bbc.com/mundo/noticias-54385775.
“Sexismo en la inteligencia artificial”, https://www.youtube.com/watch?v=PkBdBDwVvJs.

TUBELLA Patricia, “Los estudiantes británicos derrotan al algoritmo de Johnson”, *El País*, 18 de agosto de 2020, en https://elpais.com/internacional/2020-08-18/los-estudiantes-britanicos-propician-la-primera-gran-derrota-de-johnson.html.
UNESCO, “Recomendación sobre la ética de la inteligencia artificial”, 22 de noviembre de 2022, en https://unesdoc.unesco.org/ark:/48223/pf0000379920_spa.page=15.
VOSS Axel, “Draft Report on artificial intelligence in a digital age”, *European Parliament*, noviembre de 2021, en https://www.europarl.europa.eu/doceo/document/AIDA-PR-680928_EN.pdf.
WEIDINGER Laura, MELLOR John, RAUH Maribeth, GRIFFIN Conor, UESATO Jonathan, HUANG Po-Sen, HENG Myra, GLAESE Mia, BALLE Borja, KASIRZADEH Atoosa, KENTON Zac, BROWN Sasha, HAWKINS Will, STEPLETON Tom, BILES Courtney, BIRHANE Abeba, HAAS Julia, RIMELL Laura, HENDRICKS Lisa Anne, ISAAC William, LEGASSICK Sean, IRVING Geoffrey, GABRIEL Iason, “Ethical and social risks of harm from Language Models”, 8 de diciembre de 2021, p. 1-64, en https://www.deepmind.com/publications/ethical-and-social-risks-of-harm-from-language-models
WORLD HEALTH ORGANIZATION, “Ageism in artificial intelligence for health: WHO policy brief”, 2022, en https://www.analyticsinsight.net/ageism-in-artificial-intelligence-here-are-the-ways-to-prevent-,it/#:~:text=Ageism%20in%20AI%20is%20one%20new%20dimension%20to,AI%20and%20maximize%20AI%E2%80%99s%20benefits%20for%20older%20people.
XENIDIS Raphaële, “Tuning EU Equality Law to Algorithmic Discrimination: Three Pathways to Resilience”, *Maastricht Journal of European and Comparative Law 27*, n.º 6, diciembre de 2020, p. 736-758, en https://journals.sagepub.com/doi/10.1177/1023263X20982173.
XENIDIS Raphaële y SENDEN Linda, “EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination”, 30 de septiembre de 2019, p. 151-182, en https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3529524.
ZUIDERVEEN BORGESIUS Frederik, *Discrimination, artificial intelligence and algorithmic decision-making*, Directorate General of Democracy Council of Europe, 2018, en https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73.

Citar: elDial DC3045

copyright © 1997 - 2022 Editorial Albrematica S.A. - Tucumán 1440 (CP 1050) - Ciudad Autónoma de Buenos Aires - Argentina