

INFORMÁTICA Y DERECHO

REVISTA IBEROAMERICANA DE DERECHO INFORMÁTICO
(SEGUNDA ÉPOCA)

FEDERACIÓN IBEROAMERICANA DE ASOCIACIONES
DE DERECHO E INFORMÁTICA

ISSN 2530-4496 – N.º 15 VOL. 2, 2024, PP. 159-170

LOS SESGOS ALGORÍTMICOS EN LA TOMA DE DECISIONES AUTOMATIZADAS: RETOS Y OPORTUNIDADES PARA EL SISTEMA JURÍDICO PERUANO¹

*ALGORITHMIC BIAS IN AUTOMATED DECISION
MAKING: CHALLENGES AND OPPORTUNITIES
FOR THE PERUVIAN LEGAL SYSTEM*

Ángel Coaquira Flores

Universidad La Salle (Arequipa, Perú)

¹ Este trabajo recibió el segundo premio en la modalidad profesional del VIII Premio Valentín Carrascosa (2024).

Resumen

Las decisiones automatizadas mediante inteligencia artificial a primera vista son consideradas objetivas y neutrales, sin embargo, un sector opositor afirma todo lo contrario, sosteniendo que someter decisiones críticas a la soberanía de los algoritmos no solo menoscabaría habilidades humanas como el pensamiento crítico o la capacidad de resolver problemas de manera creativa, sino que también hace viable la materialización de algoritmos discriminatorios, cuya presencia en ámbitos vitales, como el de la contratación o despido del recurso humano, la generación de sentencias penales o la admisión a instituciones educativas, gestiona un estado de indefensión para un cúmulo ilimitado de derechos.

A la luz de los diversos casos en los que se ha podido evidenciar el peligro de los algoritmos discriminatorios, el análisis de la arquitectura de los algoritmos, y la evaluación del universo de dispositivos legales existentes que buscan dar respuesta a esta problemática, se pretende esbozar una propuesta en base a la realidad del sistema jurídico peruano

Palabras clave

Decisiones automatizadas, algoritmos, sesgos, inteligencia artificial, discriminación

Abstract

Automated decisions through artificial intelligence are considered objective and neutral. However, an opposing sector claims that subjecting critical decisions to the sovereignty of algorithms would not only undermine human skills such as critical thinking or the ability to solve problems creatively, but also allow the materialization of discriminatory algorithms, whose presence in vital areas (such as hiring or dismissing employees, generating criminal sentences, or admitting students in educational institutions) can cause a state of defenselessness for an unlimited number of rights.

This paper aims at outlining a proposal based on the reality of the Peruvian legal system, in light of the various cases in which the danger of discriminatory algorithms has been evident, the analysis of the architecture of the algorithms, and the evaluation of the universe of existing legal devices that seek to respond to this problem.

Keywords

Automated decisions, algorithms, biases, artificial intelligence, discrimination

Introducción

La inferencia de la inteligencia artificial en absolutamente todas las esferas sobre las que el ser humano despliega sus actividades ha generado una dependencia hacia los sistemas inteligentes, los cuales no solo se encuentran presentes en aspectos triviales, como la generación de perfiles en Spotify a través del uso de algoritmos o la creación de NPC (*Non-Player Character*) en los videojuegos mediante el uso de árboles de decisión, sino que se halla inmersa en el proceso de toma de decisiones que juegan un papel fundamental en la vida de los sujetos de derecho, como la contratación o despido del talento humano, la predicción en torno a la potencial reincidencia de los delincuentes, entre otros.

Si bien para un sector optimista el uso de algoritmos predictivos se configuraría como una forma de eliminar sesgos y, de este modo, eludir las decisiones discriminatorias debido a la pseudoneutralidad y objetividad inherente a estos sistemas, se ha evidenciado en diversos escenarios que los algoritmos han promovido el desarrollo de decisiones que atentaron con grupos sociales, debido principalmente a la caja negra de la inteligencia artificial, el proceso de extracción de data y su automaticidad y diseño productivo.

Consecuentemente, diversos países desarrollados en vista de tal desafío disruptivo, aunaron esfuerzos para diseñar o actualizar instrumentos legales; no obstante, la ineficacia de diversos cuerpos normativos y la evolución exponencial de los dispositivos inteligentes, ponen en evidencia nuevamente el divorcio entre el derecho y la realidad, hecho que presenta mayor repercusión en países en vías de desarrollo, los cuales probablemente debido al desconocimiento no ven por necesario regular el uso y abuso de la inteligencia artificial.

De este modo, se hace imperioso analizar la causa y las consecuencias de las decisiones discriminatorias promovidas por IA, así como evaluar la forma en la que diversos países abordan la problemática, para que finalmente, se pueda esbozar una propuesta para intentar regular este tipo de sesgos desde el enfoque peruano.

Las decisiones en la era de la inteligencia artificial

La toma de decisiones se configura como aquel proceso a través del cual un sujeto, organización o grupo elige entre dos o más acciones posibles dentro de una situación específica². Cabe resaltar que este proceso es considerado como una de las actividades más comunes del ser humano, es más, las personas se encuentran realizando esta tarea constantemente sobre un cúmulo de temas diferentes³.

2 Tenenbaum, G. «Decision Making in Sport», en *Encyclopedia of Applied Psychology* (Ed.: C. Spielberger), Academic Press, 2020, pp. 575-584. <https://doi.org/10.1016/B0-12-657410-3/00802-3>

3 Mendoza, M. y Gutiérrez-Peña, E., «Decision Theory», en *International Encyclopedia of Education* (Eds.: P. Peterson, E. Baker y B. McGaw), Elsevier Science, 2010, pp. 119-124.

En esa misma línea, Attwood *et al.*⁴ sostienen que aquellos sujetos que deben tomar decisiones con celeridad tienen que sopesar la velocidad en el proceso de la toma de decisiones por la precisión del producto de la decisión; asimismo, los autores indican que el tiempo requerido para tomar una decisión precisa se encuentra relacionado directamente con el grado de incertidumbre en la decisión, es decir, un mayor grado de incertidumbre demanda mayor tiempo para: la búsqueda de información, la estimación en torno a las probabilidades de las diferentes alternativas y la asignación de los valores a los resultados. Es en ese sentido que se origina un bagaje de alternativas para la mejora de las decisiones humanas, dentro de las que se encuentran: *el diseño o rediseño de sistemas de soporte*, los que estarían enfocados en mejorar tanto las decisiones como el rendimiento; y *las herramientas de decisión*, las cuales harían viable la eliminación de un determinado número de pasos que los operadores deben seguir para tomar una decisión, conviene aclarar que lo presente se puede llevar a cabo gracias a los algoritmos predictivos de aprendizaje automático⁵.

Es así que las aparentes bondades promovidas por los sistemas automatizados de toma de decisiones, como la posibilidad de generar decisiones más objetivas, neutrales y eficientes que los humanos —los cuales son susceptibles a los conflictos de interés, prejuicios o fatiga—, hacen que cada vez sean más las instituciones pertenecientes a diversos sectores que se suman a confiar en las decisiones con el apoyo de algoritmos o el actuar sobre decisiones automatizadas en las cuales no existió participación humana alguna⁶. Sin embargo, un sector crítico afirmarí­a que esta dependencia humano-máquina en la toma de decisiones conllevarí­a a un conglomerado de problemáticas, tales como la desaparición gradual de las capacidades mentales propias del ser humano como el pensamiento crítico, el análisis intuitivo y la resolución creativa de problemas⁷; el despliegue de decisiones sesgadas y discriminatorias y otra serie de cuestiones éticas⁸.

Así, creemos que si bien los sistemas de decisiones automatizadas han penetrado campos banales mediante la creación de perfiles para aplicaciones como Spotify y YouTube o la personalización de publicidad basada en el comportamiento de los cibernautas; existen decisiones basadas en algoritmos que tienen

4 Attwood, D., Deeb, J., Danz-Reece, M., «Personal Factors», en *Ergonomic Solutions for the Process Industries* (Eds.: D. Attwood, J. Deeb y M. Danz-Reece), Gulf Professional Publishing, 2004, pp. 29-63.

5 Attwood, Deeb, Danz-Reece, «Personal Factors...».

6 Koutsikouri, D., Hylving, L., Bornemark, J. y Lindberg, S., «Human judgment in the age of automated decision-making systems», en *Research Handbook on Artificial Intelligence and Decision Making in Organizations* (Eds.: I. Constantiou, M. Joshi y M. Stelmazak), Edward Elgar Publishing, 2024, p. 144-159.

7 Fayaz Ahmad, S., Han, H., Mansoor, M., Khairul, M., Irshad, M., Arraño-Muñoz, M., Ariza-Montes, A., «Impact of artificial intelligence on human loss in decision making, laziness and safety in education», *Humanities and Social Sciences Communications*, 10, 2023.

8 Wirtz, B., Weyerer, J., Sturm, B., «The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public Administration», *International Journal of Public Administration*, vol. 43, n.º 9, 2020, pp. 818-829.

un impacto significativo en la vida de todos los seres humanos, ya sea para la contratación o despido en el personal de una entidad, la generación de sentencias penales o la evaluación de elegibilidad para la liberación anticipada de un reo⁹, por ende, la discriminación algorítmica basada en sesgos u otros factores aunado a la no regulación, sitúan en un estado de indefensión a un cúmulo considerable de derechos. Consecuentemente, analizaremos el tratamiento de la discriminación desde diversas ópticas para que posteriormente se estudie la colisión existente entre la marginación y la inteligencia artificial.

La evolución de la discriminación: hacia la exclusión algorítmica

Todo ser humano, durante la niñez, aprende un conjunto de tradiciones y normas propias de la sociedad en la que nació, es de menester indicar que tanto los prejuicios como la discriminación son parte de esas normas y tradiciones, las cuales forman parte de nuestro patrimonio cultural; y si bien esta parte del patrimonio no suele ser aceptado por muchas personas, nadie puede escapar a lidiar con este¹⁰. Además, el autor indica que los psicólogos sociales hallaron que la edad entre la que los niños suelen aprender los prejuicios grupales oscila entre los tres y cuatro años, incluso antes de siquiera tener la capacidad de distinguir entre grupos, es así que con el discurrir del tiempo se hace necesario que el menor desaprenda este tipo de actitudes y normas. No obstante, estas actitudes, al formar parte de la cultura, se preparan constantemente hasta que alcanzan a convertirse en filtros mediante los cuales eventualmente observamos y clasificamos a sujetos y grupos.

En esa misma línea, dentro de los estudios a través de los que se analizan las discriminaciones llevadas a cabo en instituciones, sobresale el de Michael Schaefer *et al.*¹¹, quienes mediante un meta-análisis con base en 85 auditorías de campo y 361.645 solicitudes de empleo personales probaron el sesgo de género durante la etapa de contratación en trabajos estereotípicos, tanto para hombres como mujeres, entre 1976 y 2020. Además, los autores identificaron que aun cuando hasta hace poco se observaba una preferencia respecto al género masculino durante los procesos de selección y contratación para muchos cargos, este tipo de discriminación ha disminuido en algunos países desarrollados, llegando a generarse una ligera inversión a favor de las candidatas a puestos de trabajo. Por otro lado, Pager y Shepherd¹² reconocen que la desigualdad racial persiste en ámbitos como la vivienda o el empleo, sin embargo, si bien antes de la era

- 9 Araujo, T., Helberger, N., Kruikemeier, S. De Vreese, C., «In AI we trust? Perceptions about automated decision-making by artificial intelligence», *AI & Society*, vol. 35, 2020, pp. 611-623.
- 10 Ehrlich, H. *Hate Crimes and Ethnoviolence: The History, Current Affairs, and Future of Discrimination in America*. Westview Press, 2009.
- 11 Schaefer, M. *et al.*, «On the trajectory of discrimination: A meta-analysis and forecasting survey capturing 44 years of field experiments on gender and hiring decisions», *Organizational Behavior and Human Decision Processes*, vol. 179, 2023.
- 12 Pager, D., Shepherd, H., «The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets», *Annual Review of Sociology*, vol. 1, n.º 34, 2008, pp. 181-209.

de los derechos civiles el racismo era evidente y generalizado, hoy en día se evidencia que la discriminación es cada vez más difícil de identificar, situación que implica problemas de medición y conceptualización para las ciencias sociales.

Dentro de este orden de ideas, se puede apreciar que los sesgos y la discriminación forman parte del patrimonio humano; de la misma manera, el tiempo y la intromisión de los derechos han hecho que la tarea de identificar actos discriminatorios sea cada vez menos viable. Ahora bien, a pesar de que un vasto número de personas puede considerar a la discriminación desarrollada por máquinas como un suceso que solo puede darse en la ciencia ficción, al revisar el gran número de casos que involucran algoritmos excluyentes no solo se pone en evidencia la palpable problemática objeto de la investigación, sino que, además encontramos que históricamente, los primeros casos de marginación promovidos por la tecnología se originan desde el siglo XX.

Es así que, a inicios de la década de los ochenta, en Reino Unido la escuela médica del Hospital St. George se vio obligada a crear un programa informático de selección de postulantes, en vista de la gran demanda de estudiantes para ingresar a sus claustros y la necesidad de reducir la carga laboral durante la selección de candidatos para la entrevista¹³. Posteriormente, en 1986, los catedráticos A. Burke y J. Collier informaron a la comisión universitaria que el programa empleado en la selección inicial de solicitantes para la escuela discriminaba de manera injusta a las mujeres y a aquellas personas cuyos nombres no sonaran europeos¹⁴.

Asimismo, basándonos en la clasificación de discriminación algorítmica propuesta por Chen¹⁵, analizaremos los casos más resaltantes de exclusión hechas por IA:

- Discriminación por género: O'Connor y Liu¹⁶ establecen que este tipo de exclusión se puede materializar en: *la traducción automática*, debido al sesgo que posee la IA hacia los valores predeterminados masculinos, sobre todo en campos como STEM ya que mantienen una inclinación hacia un género; *las herramientas de reconocimiento facial*, en razón de que se pudo evidenciar que los puntos de referencia, los cuales en términos de visión por computadora son un conjunto de imágenes que muestran rostros de personas de todo el espectro de género y raza, sobre presentaban a los individuos masculinos y de piel clara, mientras que subpresentaban a las personas de piel más oscura.
- Discriminación por carrera: el *chatbot* de Microsoft, Tay, producto de la interacción con cibernautas en la plataforma y la data para entrenamiento

13 Barocas, S., Selbst, A., «Big Data's Disparate Impact», *California Law Review*, 104, 2016, pp. 671-732

14 Lowry, E., Macpherson, G., «A blot on the profession», *British Medical Journal*, 5 de marzo, 1988, pp. 657-658.

15 Chen, Z., «Ethics and discrimination in artificial intelligence-enabled recruitment practices», *Humanities and social sciences communications*, 10, 2023.

16 O'Connor, S., Liu, H., «Gender bias perpetuation and mitigation in AI technologies: challenges and opportunities», *AI & Society*, 2023.

proveniente de *tweets* humanos, aprendió a generar comentarios sexistas y racistas, de hecho, el «inocente» *bot* adoptó un discurso de odio hacia las mujeres y personas de color; como consecuencia de ello, un sector afirma que las máquinas son capaces de absorber sesgos humanos de manera pasiva y reproducirlos de manera inconsciente¹⁷.

- Discriminación por color de piel: la aplicación de *software* de Gestión de Perfiles de Delincentes Correccionales para Sanciones Alternativas (COM-PAS) se emplea para informar las decisiones sobre libertad condicional y libertad bajo fianza¹⁸; sin embargo, la organización de ProPublica en el 2016 investigó la IA, llegando a concluir que existía una alta probabilidad de que el algoritmo posea sesgos contra los acusados afroamericanos, ya que tenían más probabilidades de ser etiquetados con el mayor riesgo incluso si no poseían condenas previas¹⁹. Aunado a ello, encontramos la aplicación de fotografía de Google que etiquetó una foto de dos personas de color como gorilas en el 2015²⁰.
- Discriminación por personalidad: si bien esta clase de discriminación sería innovadora para diversos sistemas jurídicos, llama la atención el caso *Mobley v. Workday, Inc.*, el cual versa sobre un hombre afroamericano de 40 años, con una licenciatura en finanzas y un posgrado con honores, quien alega que sufre de depresión y ansiedad y que, desde el 2017, solicitó trabajo a más de cien empresas, las cuales emplean Workday, una herramienta de selección de recursos humanos que realiza su tarea mediante la aplicación de una prueba de personalidad, y absolutamente todas las solicitudes tuvieron una respuesta negativa. Asimismo, el demandante indica que incluso en una oportunidad solicitó trabajo a las 12:55 a. m., obteniendo el rechazo del sistema una hora después. Consecuentemente, el demandante sostiene que el algoritmo posee un sesgo respecto a los usuarios que posean trastornos de salud mental o deterioro cognitivo, tengan procedencia afroamericana, posean más de 40 años o sean discapacitados²¹. Cabe señalar que el tribunal dio amparo a los reclamos de discriminación por impacto dispar.

En suma, se advierte del análisis de casos previo que la exclusión algorítmica atenta contra la dignidad *per se* y diversos derechos fundamentales de los seres humanos, lo cual obliga a los Estados no solo a generar un marco regulatorio *ad hoc*, sino que además, en vista de la evolución constante de este tipo de tecnologías, se debe considerar formas de dinamizar el derecho sin la necesidad de enfrentarse a barreras burocráticas. Así, sostenemos que en primer lugar debemos analizar la causa de los problemas algorítmicos y posteriormente evaluar

17 O'Connor y Liu, «Gender bias...».

18 Huq, A., «Racial Equity in Algorithmic Criminal Justice». *Duke Law Journal*, vol. 68, n.º 6, 2019, pp. 1043-1134.

19 Ferrara, E. «Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources», *Impacts, and Mitigation Strategies. Sci*, vol. 6, n.º 1, 2024.

20 Chen, «Ethics and discrimination...».

21 Karasik, A., Maatman, G., Schaller, G., «California Federal Court Denies Motion To Dismiss Artificial Intelligence Employment Discrimination Lawsuit. Class Action Defense», *DuaneMorris.com* (blog), 16 de julio, 2024.

los dispositivos normativos propuestos hasta el momento para que, de esta forma, se esboce un marco basado en la realidad peruana.

Desmantelando al algoritmo: identificando las causas de la discriminación 2.0

De acuerdo con Cossette Lefebvre y Maclure²², lo preocupante de la discriminación generada por algoritmos predictivos de aprendizaje automático es que este acto excluyente surge de las características inherentes a este tipo de algoritmos, las cuales son:

- Discriminación por extracción de datos y categorización, aspecto que se materializa en cuatro formas:

1) *La distinción entre variable objetivo y etiqueta de clase puede introducir determinados sesgos en el funcionamiento del algoritmo*, considerando a la variable objetivo como los resultados de interés, y las etiquetas como aquellos elementos encargados de dividir el valor de las variables objetivo en categorías excluyentes; el autor plantearía como ejemplo: aquel algoritmo que debe decidir si un correo electrónico es *spam* (variable objetivo), para lo cual se basa en dos etiquetas de clase: un correo es *spam* o no, estableciéndose distinciones entre ambas etiquetas; siendo así, el problema se ocasiona al momento de definir la variable objetivo, lo cual es una empresa difícil de ejecutar, y se aprecia en ejemplos de variables como «buen alumno» o «criminal potencialmente peligroso». En ese sentido, la misma definición puede generar una discriminación de impacto dispar.

2) *El peligro de que los algoritmos de machine learning son muy inductivos*, ya que el programa se entrena en base a un conjunto de ejemplos, siendo así, es altamente probable que los ejemplos empleados introduzcan sesgos al algoritmo.

3) *Debido a la imposibilidad de poder capturar todos los aspectos de los objetos en general, las entidades se ven obligadas a tomar decisiones sobre lo que pueden observar y posteriormente añadir un análisis*, el autor ejemplifica lo presente mediante el algoritmo que adjudica un gran peso a la reputación de la Universidad de un postulante, variable que no necesariamente demostrara las habilidades del potencial ingresante, hecho que perjudica a otros grupos marginados.

4) *Los sustitutos elegidos por los programadores*, lo cual colisiona con diversas normas relacionadas a la protección de datos personales, en vista de que existen datos sensibles que no pueden ser empleados para el entrenamiento de las máquinas inteligentes, los tecnólogos se ven obligados a emplear datos sustitutos.

- Discriminación mediante la automaticidad: durante el proceso predictivo encontramos la disyuntiva respecto al evaluar si es que emplear correlaciones observadas en un grupo para guiar la toma de decisiones de un individuo es

22 Cossette-Lefebvre, H., Maclure, J., «AI's fairness problem: understanding wrongful discrimination in the context of automated decision-making», *AI and Ethics*, vol. 3, 2023.

discriminatorio, ya que se corre el riesgo de utilizar generalizaciones erróneas para dictaminar casos particulares.

- Discriminación y opacidad: quizás la característica más polémica es el llamado «algoritmo de caja negra» el cual se fundamenta en la opacidad inherente de los sistemas inteligentes, ya que, si bien es posible explicar la arquitectura de un algoritmo y de esta forma aislar las variables predictivas que emplea ensayando con su actuar, nunca se podrá tener conocimiento sobre cómo es que los algoritmos alcanzan un resultado en particular; es más, si se busca una mayor precisión en el resultado el costo de ello es una mayor opacidad.

En síntesis, se constata que el algoritmo es potencialmente discriminatorio en razón de su misma naturaleza; y es que, tanto su arquitectura como mecanismo de funcionamiento turbio generan un cúmulo ilimitado de problemáticas de carácter ético. Siendo así, en aras de esbozar una propuesta para regular los algoritmos en la realidad peruana, se observará diversos dispositivos legales confeccionados por otros países.

Dispositivos legales *versus* sesgos algorítmicos: ¿la cura resulta más mala que la enfermedad?

Frente a las cuestiones promovidas por los algoritmos excluyentes, se generaron una constelación de propuestas y remodelaciones normativas en diversos países, es por esto que para una mejor comprensión se clasificaran las iniciativas en dos bloques, *hard law* y *soft law*. Es así que, dentro del primer bloque, *hard law*, encontramos que la discriminación se encuentra prohibida en diversos tratados internacionales, constituciones o códigos penales de Latinoamérica, sin embargo, se hace necesario el analizar el artículo 14 del Convenio Europeo de los Derechos Humanos, del cual se desprende dos tipos de discriminación, la directa e indirecta; concibiendo a la primera como la posibilidad de que las personas sean discriminadas con base en una característica protegida normativamente, y a la segunda como aquella práctica que a primera vista puede reflejar neutralidad, sin embargo, se termina discriminando a una persona de un determinado origen racial u otro impacto protegido²³. Es así que podemos advertir que la IA únicamente podría desarrollar la discriminación secundaria, no obstante, los regímenes antidiscriminatorios presentarían deficiencias al tratar de regular los algoritmos excluyentes, ya que, según Prince y Schwarcz²⁴, los sistemas inteligentes hacen viable que: 1) se trasladen los costos provenientes de las características predictivas al grupo protegido; 2) se permita que el discriminador se aproveche indirectamente del poder predictivo de rasgos sospechosos; 3) se reproduzca y refuerce los legados de discriminación histórica basándose en una justificación económica para los discriminadores; y 4) se genere la discriminación por parte de la carga actuariamente a los elementos de un grupo protegido con las características generales de su grupo.

²³ Zuiderveen, F., *Discrimination, Artificial Intelligence, and Algorithmic Decision Making*. Directorate General of Democracy, 2018.

²⁴ Prince, A., Schwarcz, D., «Proxy Discrimination in the Age of Artificial Intelligence and Big Data», *Iowa Law Review*, vol. 105, n.º 3, 2020.

En esa misma línea, encontramos a toda la normativa internacional que se enfoca en la protección de datos personales, sobre todo la de la Unión Europea, es decir, el *Reglamento General de Protección de Datos* (RGPD), sobre el cual se identifican dos aspectos fundamentales: el artículo 22, denominado «Decisiones individuales automatizadas, incluida la elaboración de perfiles», a través del cual se prohíbe las decisiones totalmente automatizadas que promuevan efectos significativos y, por ende, salvaguarda los derechos de los ciudadanos a verse afectados por este tipo de decisiones en todas sus aristas; y las diversas normativas que buscan proteger y restringir el tratamiento de data sensible en general, sin embargo, esta tutela puede configurarse como un arma de doble filo en el sentido de que los programadores al verse prohibidos de emplear este tipo de data buscaran sustitutos, los cuales finalmente lleven a una potencial discriminación.

Por otro lado, respecto a las herramientas de *soft law* para abordar a los algoritmos excluyentes hallamos la promoción del buen gobierno corporativo o las sobreestimadas auditorías de inteligencia artificial, las cuales en palabras de Van Beekum y Zuiderveen²⁵ encontrarían dos obstáculos: la necesidad de utilizar categorías especiales de datos —lo cual está prohibido por ley— para la ejecución de pruebas a la IA contra la discriminación, y la cada vez más inviable posibilidad de materializar una auditoria eficaz debido a que muchos profesionales no cuentan con la infraestructura para recopilar categorías especiales de datos y al hecho de que los métodos no son lo suficientemente holísticos.

En suma, el hecho de que Perú sea considerado como uno de los países con mayor diversidad étnica y lingüística²⁶ hace que el Estado se vea obligado a desarrollar un cuerpo normativo que abarque tanto la discriminación directa como la indirecta, a pesar de que se puede evidenciar actualmente dispositivos que regulan la discriminación en el ámbito laboral y administrativo; del mismo modo, se debe gestionar la renovación de la ley de protección de datos personales, a través de la cual se prohíba el producto de las decisiones automatizadas en donde no hubo participación humana alguna y, además, redimensionar la naturaleza de los datos sensibles para hacer viable las auditorias algorítmicas; finalmente, creemos que la ley de protección de datos debe promover un principio que obligue a los programadores a demostrar la arquitectura y el funcionamiento del algoritmo que construyen: el principio de la inteligencia artificial imparcial.

Consideraciones finales

Creemos que, debido a las características inherentes al algoritmo, como la caja negra, y la carencia de normas para regular los sistemas inteligentes de manera efectiva, se ha configurado un escenario mundial en el cual absolutamente

25 Van Beekum, M., Zuiderveen, F., «Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception?», *Computer Law & Security Review*, vol. 48, 2020.

26 Mejía, D., Manayay, E., Valenzuela, G., Benavides, H., Hilario, V., *Perú: Estado de la Población en el año del Bicentenario, 2021*, Instituto Nacional de Estadística e Informática, 2021.

todas las personas son pasibles de ser discriminadas por un algoritmo, incluso sin darse cuenta, lo que es más llamativo.

Ahora bien, sostenemos que la forma óptima de tratar la problemática algorítmica no es solo gestionar un cúmulo de normas *ad hoc*, sino que también se debe buscar aunar los esfuerzos de diversas instituciones privadas y públicas —tales como ONG que luchan contra la discriminación, ministerios del Estado que promuevan la tecnología o resguarden a los grupos minoritarios, empresas en general— para atenuar los efectos de algoritmos excluyentes mediante la gestión de políticas, normativas institucionales, entre otros.

Referencias bibliográficas

- Araujo, T., Helberger, N., Kruikemeier, S. De Vreese, C. «In AI we trust? Perceptions about automated decision-making by artificial intelligence», *AI & Society*, vol. 35, 2020, pp. 611-623. <https://doi.org/10.1007/s00146-019-00931-w>
- Attwood, D., Deeb, J., Danz-Reece, M. «Personal Factors», en *Ergonomic Solutions for the Process Industries* (Eds.: D. Attwood, J. Deeb y M. Danz-Reece), Gulf Professional Publishing, 2004, pp. 29-63. <https://doi.org/10.1016/B978-075067704-2/50004-0>
- Barocas, S., Selbst, A. «Big Data's Disparate Impact», *California Law Review*, 104, 2016, pp. 671-732. <http://dx.doi.org/10.2139/ssrn.2477899>
- Chen, Z. «Ethics and discrimination in artificial intelligence-enabled recruitment practices», *Humanities and social sciences communications*, 10, 2023. <https://doi.org/10.1057/s41599-023-02079-x>
- Cossette-Lefebvre, H., Maclure, J. «AI's fairness problem: understanding wrongful discrimination in the context of automated decision-making», *AI and Ethics*, vol. 3, 2023, <https://doi.org/10.1007/s43681-022-00233-w>
- Ehrlich, H. *Hate Crimes and Ethnoviolence: The History, Current Affairs, and Future of Discrimination in America*. Westview Press, 2009.
- Fayaz Ahmad, S., Han, H., Mansoor, M., Khairul, M., Irshad, M., Arraño-Muñoz, M., Ariza-Montes, A. «Impact of artificial intelligence on human loss in decision making, laziness and safety in education», *Humanities and Social Sciences Communications*, 10, 2023. <https://doi.org/10.1057/s41599-023-01787-8>
- Ferrara, E. «Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources», *Impacts, and Mitigation Strategies. Sci*, vol. 6, n.º 1, 2024. <https://doi.org/10.3390/sci6010003>
- Huq, A. «Racial Equity in Algorithmic Criminal Justice». *Duke Law Journal*, vol. 68, n.º 6, 2019, pp. 1043-1134.
- Karasik, A., Maatman, G., Schaller, G. «California Federal Court Denies Motion To Dismiss Artificial Intelligence Employment Discrimination Lawsuit. Class Action Defense», *DuaneMorris.com* (blog), 16 de julio, 2024, <https://blogs.duanemorris.com/classactiondefense/2024/07/16/>

- california-federal-court-denies-motion-to-dismiss-artificial-intelligence-employment-discrimination-lawsuit/
- Koutsikouri, D., Hylving, L., Bornemark, J., Lindberg, S. «Human judgment in the age of automated decision-making systems», en *Research Handbook on Artificial Intelligence and Decision Making in Organizations* (Eds.: I. Constantiou, M. Joshi y M. Stelmaszak), Edward Elgar Publishing, 2024, pp. 144-159. <http://dx.doi.org/10.4337/9781803926216>
- Lowry, E., Macpherson, G. «A blot on the profession», *British Medical Journal*, 5 de marzo, 1988, pp. 657-658. <https://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC2545288&blobtype=pdf>
- Mejía, D., Manayay, E., Valenzuela, G., Benavides, H., Hilario, V. *Perú: Estado de la Población en el año del Bicentenario, 2021*. Instituto Nacional de Estadística e Informática, 2021. https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1803/libro.pdf
- Mendoza, M., Gutiérrez-Peña, E. «Decision Theory», en *International Encyclopedia of Education* (Eds.: P. Peterson, E. Baker y B. McGaw), Elsevier Science, 2010, pp. 119-124. <https://doi.org/10.1016/B978-0-08-044894-7.01319-1>
- O'Connor, S., Liu, H. «Gender bias perpetuation and mitigation in AI technologies: challenges and opportunities», *AI & Society*, 2023. <https://doi.org/10.1007/s00146-023-01675-4>
- Pager, D., Shepherd, H. «The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets», *Annual Review of Sociology*, vol. 1, n.º 34, 2008, pp. 181-209. <https://doi.org/10.1146/annurev.soc.33.040406.131740>
- Prince, A., Schwarcz, D. «Proxy Discrimination in the Age of Artificial Intelligence and Big Data», *Iowa Law Review*, vol. 105, n.º 3, 2020.
- Schaerer, M., Du Plessis, C., Bao, M., Van Aert, R., Tiokhin, L., Lakens, D., Clemente, E., Pfeiffer, T., Dreber, A., Johannesson, M., Clark, C., Gender Audits Forecasting Collaboration, Uhlmann, E. «On the trajectory of discrimination: A meta-analysis and forecasting survey capturing 44 years of field experiments on gender and hiring decisions», *Organizational Behavior and Human Decision Processes*, vol. 179, 2023, <https://doi.org/10.1016/j.obhdp.2023.104280>
- Van Bakkum, M., Zuiderveen, F. «Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception?», *Computer Law & Security Review*, vol. 48, 2020. <https://doi.org/10.1016/j.clsr.2022.105770>
- Wirtz, B., Weyerer, J., Sturm, B. «The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public Administration», *International Journal of Public Administration*, vol. 43, n.º 9, 2020, pp. 818-829. <https://doi.org/10.1080/01900692.2020.1749851>
- Zuiderveen, F. *Discrimination, Artificial Intelligence, and Algorithmic Decision Making*. Directorate General of Democracy, 2018.