DISCRIMINACIÓN ALGORÍTMICA EN EL ÁMBITO LABORAL

ALGORITHMIC DISCRIMINATION AT WORK

Mario Eduardo MALDONADO SMITH¹

RESUMEN: El presente artículo analiza las potenciales vías de discriminación que pueden darse mediante el empleo de algoritmos informáticos, en especial aquellos que incorporan tecnología de Inteligencia Artificial. Se presentan consideraciones generales sobre esta tecnología y luego se plantean escenarios de discriminación que pueden darse en lo general y específicamente en el ámbito laboral. El objetivo es presentar diversos aspectos que habrían de ser cuidados en el empleo de esta tecnología de manera que puedan resultar armonizados los diversos intereses en juego. Tras analizar cómo nuestro ordenamiento jurídico puede responder a este fenómeno, se concluye con una serie de principios y “buenas prácticas” que podrían ser incorporadas en nuestra legislación. Como métodos de trabajo se han empleado el comparado, analítico, jurídico e inferencial.

PALABRAS CLAVE: Algoritmo, Inteligencia Artificial, discriminación, derechos humanos, trabajo.

ABSTRACT: *This article analyzes the potential ways discrimination can occur through the use of computer algorithms, especially those incorporating Artificial Intelligence technology. General considerations about this technology are presented, followed by scenarios of discrimination that may occur in general and specifically at work. The aim is to present various aspects that should be taken care of when employing this technology so that the different interests at stake can be harmonized. After analyzing how our legal system can respond to this phenomenon, the article concludes with a series of principles and “best practices” that could be incorporated into our legislation. Comparative, analytical, legal, and inferential methods have been used in this work.*

KEYWORDS: *Algorithm, Artificial Intelligence, Discrimination, Human Rights, Work.*

¹ Investigador A del Centro de Estudios de Derecho e Investigaciones Parlamentarias de la Cámara de Diputados y profesor de la Universidad Internacional de La Rioja, Doctor en Derechos Humanos por la Università degli Studi di Napoli, Federico II.

SUMARIO: I. *Introducción. Revolución industrial 4.0 y su impacto en el ámbito laboral.* II. *Consideraciones generales sobre la igualdad y la no discriminación.* III. *La discriminación algorítmica en el ámbito laboral.* IV. *Marco normativo contra la discriminación laboral algorítmica.* V. *Reflexiones desde la óptica legislativa. Retos y oportunidades.* VI. *Conclusiones.* VII. *Bibliografía.*

I. INTRODUCCIÓN. REVOLUCIÓN INDUSTRIAL 4.0 Y SU IMPACTO EN EL ÁMBITO LABORAL

El descubrimiento o la invención de una determinada técnica, habilidad o tecnología ha dado paso a nuevas divisiones y categorizaciones en el trabajo, tal como se atestigua con el descubrimiento del fuego o la agricultura, así como con la invención de la escritura, la imprenta o la máquina de vapor.²

Las nuevas tecnologías suponen nuevas formas de división del trabajo en las que también se presentan nuevos temores en tanto que, cada una de esas tecnologías plantea el tránsito o desplazo de la anterior. Estos procesos no siempre son apacibles, tal como ocurrió en la Revolución industrial, donde la presencia de la máquina de vapor y sus efectos (reducción de personal) fue vista por algunos (los gremios) como una amenaza y, por tanto, como un enemigo a destruir.³

Hoy en día nos encontramos ante una tecnología que con mucho atino es ya considerada por algunos como una nueva Revolución Industrial (la llamada “Revolución 4.0”),⁴ que se refiere a la incorporación de la Inteligencia Artificial (IA) en los procesos productivos.⁵ En efecto, aunque

² Rifkin, *La sociedad del coste marginal cero. El internet de las cosas, el procomún colaborativo y el eclipse del capitalismo*, 49-51.

³ Huberman, *Los bienes terrenales del hombre*, 208-212.

⁴ Rifkin, *La sociedad del coste...*, 23; Ramírez Aufrán, “Sesgos y discriminaciones sociales de los algoritmos en Inteligencia Artificial: una revisión documental”, 2.

⁵ De acuerdo con el Reglamento de Inteligencia Artificial de la Unión Europea, un sistema de IA puede ser definido como: “Un sistema basado en una máquina diseñado para funcionar con distintos niveles de autonomía, que puede mostrar capacidad de adaptación tras el despliegue y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar información de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales”, Art. 3, numeral 1, Reglamento de Inteligencia Artificial, 13 de marzo de 2024, https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_ES.pdf (acceso el 17 de junio de 2024).

muchas veces no seamos conscientes de ello, la IA está hoy presente en prácticamente todos los aspectos de la vida y todo apunta a que con el paso de los años su avance será exponencial (en medicina, educación, política, derecho, agricultura, ganadería, finanzas, etc.).⁶

Asistimos a una verdadera revolución en la información, en la recopilación, almacenamiento, procesamiento y análisis de ingentes cantidades de datos para una sociedad ávida de consumo y de inmediatez en sus respuestas; una sociedad que, como señala Abadías Selma, se caracteriza por la reducción del tiempo y el espacio, de un “aquí y ahora”.⁷ En este escenario consumista y de respuestas inmediatas, consentimos que internet nos “recomiende” una película, un servicio, una determinada lectura, o, en el ámbito laboral, el envío de información sobre determinadas ofertas de trabajo. En todos estos casos, observaremos que, muy curiosamente, toda esa información coincide con nuestros gustos e intereses.

Desde ahora adelantamos que no es ninguna casualidad, sino el resultado de un previo análisis de los sistemas informáticos que permanentemente nos observan, miden, cuantifican, categorizan y predicen nuestro comportamiento.⁸

Sin embargo, estas sugerencias que parecerían inofensivas no siempre lo son, porque al igual que a cada uno de nosotros en lo individual, también se le presentan a grandes empresas o inclusive al Estado en cuestiones como podrían ser: el otorgamiento de créditos sociales, becas, subsidios, acceso a una determinada prestación o, inclusive, decisiones referidas a la propia libertad individual.⁹

El ámbito laboral no ha escapado a los efectos de esta nueva tecnología y, como ha ocurrido con las anteriores revoluciones industriales, su incorporación no ha estado exenta de críticas, miedos y temores por muchas y complejas razones. Como ocurre con los ejemplos que hemos

6 Soria Olivas *et. al.*, *Inteligencia Artificial. Casos prácticos con aprendizaje profundo*, 44; Molina, “Regulación de la Inteligencia Artificial aplicada a la medicina: actualidad y propuestas”, 58.

7 Abadías Selma, *Justicia juvenil e inteligencia artificial en la era de la cultura “Touch”*, 27.

8 O’Neil, *Weapons of Math Destruction. How Big Data increases inequality and threatens democracy*, 70.

9 Marcazzolo Awad y Walker Silva, “Empleo de sistemas algorítmicos de evaluación de riesgos en materia penal. Estándares mínimos para un uso acorde a las exigencias del debido proceso”, 41. La referencia a la libertad alude al empleo de algoritmos informáticos para elaborar perfiles delincuenciales sobre determinada categoría de sujetos, tecnología que incluso es empleada por algunos jueces para justificar una determinada pena o sanción, el ejemplo paradigmático es el programa “COMPAS” empleado en los Estados Unidos, véase: Caterini, “El sistema penal en la encrucijada ante el reto de la inteligencia artificial”, 6.

presentado, en la actualidad no son pocas las decisiones que se toman con base en la predicción de sistemas informáticos, ello supone que, en el contexto de la cuarta Revolución Industrial, asistamos a un momento en el que hemos dejado en manos de computadoras decisiones trascendentales, inclusive aquellas que pueden tener incidencia en la tutela y garantía de los derechos humanos.

El escenario se presenta complejo, aunque no tendría que ser así. La tecnología no es ni buena ni mala, sino funcional a los objetivos que el ser humano pueda darle.¹⁰ Así como existen críticas, miedos y temores, también se plantean innumerables espacios de oportunidad. Hay quienes afirman que la incursión de esta tecnología puede dar lugar a satisfacer las necesidades más básicas de la humanidad, lo cual incluye una vida digna sin hambre, violencia, sufrimiento, dolor o enfermedad.¹¹

Otro de los beneficios que se alegan es el de la toma de decisiones objetivas, precisas y eficientes sustentadas en la evidencia empírica y, en consecuencia, ausentes de las negligencias, subjetividades y errores humanos. Bajo este presupuesto, las decisiones trascendentales que inciden en el destino de millones de personas serían (en teoría) mejores, al ser la resultante del análisis de enormes volúmenes de datos duros, medibles y cuantificables.¹²

En el ámbito empresarial también se suele alegar que mediante esta tecnología es posible realizar de manera eficaz y eficiente un gran número de tareas que requieren elevados recursos (en tiempo, fuerza de trabajo, materia prima, etc.), de manera que la IA posibilita determinar las mejores estrategias de optimización que, a su vez, permiten redireccionar recursos para mejorar la producción. Si se reflexiona, más allá de un cierto mantenimiento, una máquina puede trabajar las 24 horas del día, no recibe un salario o vacaciones, ni tampoco tienen que pagarse gastos derivados de su incorporación a la seguridad social. Si a esta máquina se le incorpora tecnología de IA, los beneficios resultan abrumadores.

10 Rodotà, *Tecnopolitica. La democrazia e le nuove tecnologie della comunicazione*, 28.

11 Domingos, *The Master Algorithm*, 44-46; Tegmark, *Vida 3.0. Ser humano en la era de la Inteligencia Artificial*, 193; Bostrom, *Superinteligencia. Caminos, peligros, estrategias*, 236.

12 Morondo Taramundi, "Le sfide della discriminazione algoritmica", 23.

Por el lado contrario podemos reflexiones en qué es lo que ocurre con la incorporación de esta tecnología para el trabajador.¹³ En primer lugar, como también ocurrió en el pasado, actividades laborales que antaño requerían la intervención humana, ahora no la precisan y en un futuro próximo todo permite avizorar que serán más los campos en donde la presencia humana será innecesaria, especialmente tratándose de actividades susceptibles de automatización. Por citar algunos ejemplos en donde esto ya ocurre: los “Chatbots” incorporados en páginas web o sistemas de telefonía; herramientas virtuales para el pago de servicios; empresas 100% on-line; “cajas de cobro” automatizadas en centros comerciales; servicios digitales a través de “Apps” (donde incluso no hay trabajadores como tales, sino “socios” o “colaboradores independientes”); y qué decir de los modelos de lenguaje de gran tamaño (Large Language Model, LLM), como es el caso del ChatGPT, herramientas que al ser perfeccionadas podrán desplazar a ciertas profesiones dedicadas a la consulta especializada (abogados, médicos, contadores, consultores, etc.).¹⁴

Este contexto obliga a repensar la formación y desarrollo de nuevas habilidades digitales para poder seguir el rumbo de la modernidad y no quedar desplazados por la misma, pero, además, obliga a reflexionar cómo esta tecnología funciona y qué implicaciones puede tener para los derechos laborales. Como se advierte, los retos, miedos y dilemas son inmensos, en esta investigación nos concentraremos tan sólo en un fenómeno que se refiere al empleo de la tecnología algorítmica (de IA) como una eventual forma de discriminación en el ámbito laboral.

En efecto, otro de los riesgos que se ciernen sobre los derechos laborales tiene que ver con el empleo de la tecnología algorítmica al ámbito decisional empresarial. Actualmente, las grandes empresas toman sus decisiones con base en programas informáticos en los que se emplean algoritmos de IA. El mayor desarrollo tecnológico no hará sino potencializar esta situación y en un próximo futuro programas de este tipo podrán estar a disposición de cualquier empleador.

Como veremos, la decisión algorítmica puede partir de sesgos que terminan provocando tratos discriminatorios para ciertas categorías de

¹³ En una perspectiva más amplia: Maldonado Smith, “Reflexiones iuspenalistas sobre los delitos contra la libertad sindical en la era digital”, 261-264.

¹⁴ Tegmark, *Vida 3.0. Ser humano...*, 141.

sujetos. Estos sesgos pueden ser el resultado de diferentes factores, en algunos casos de la inclusión consciente de categorías como el sexo, la edad o el color de piel al momento de configurar los algoritmos, pero también el resultado de una decisión inconsciente o, inclusive, autónoma del propio programa informático. Además, se presentan dificultades probatorias que no siempre pueden superarse al momento de acreditar una eventual vulneración a los derechos fundamentales, sin contar con que, en algunos casos, esa vulneración puede ser tan sutil que sea incluso difícil de percibir. Si bien estas reflexiones pueden ser aplicadas a lo largo de toda la relación laboral, por su alto potencial discriminatorio nos referiremos ante todo a la etapa de contratación en donde se observará cómo la discriminación algorítmica puede tener lugar.

La justificación de esta investigación se asienta en la relevancia del tema dentro de la agenda legislativa de la Cámara de Diputados, en donde existen ya diversas propuestas para regular las implicaciones de la IA,¹⁵ en este sentido, busca ser una contribución al ámbito jurídico-laboral, pero, ante todo, al ámbito de la tutela de los derechos fundamentales frente a los riesgos de las nuevas tecnologías.

Como hipótesis general planteamos que el empleo de algoritmos de IA en el ámbito laboral es susceptible de vulnerar el derecho fundamental a la igualdad que, a su vez, tiene como corolario la no discriminación, principio cúspide sobre el que descansa la dignidad del ser humano.¹⁶ En este caso, una dignidad que supone no instrumentalizar a la persona, sino considerarla un fin, tal y como se desprende del artículo 1.º, párrafos segundo, tercero y quinto, de la Constitución Política de los Estados Unidos Mexicanos (CPEUM). A la par, se plantea que no sólo la igualdad y la no discriminación pueden verse afectados, sino otros tantos derechos de orden fundamental que a lo largo de esta investigación se presentan.

Como métodos de estudio se emplea el comparado, analítico, jurídico e inferencial, para ello se acude a la doctrina más especializada en la materia, a la jurisprudencia que existe en el rubro, a casos ocurridos en otros países, así como a reflexiones de orden jurídico y de principios.

¹⁵ “Cámara de Diputados impulsa acciones ante desafíos de la Inteligencia Artificial”, 27 de diciembre de 2023, acceso el 17 de junio de 2024, https://www.canaldelcongreso.gob.mx/noticias/17403/Camara_de_Diputados_impulsa_acciones_ante_desafios_de_la_Inteligencia_Artificial

¹⁶ Pinto, *Temas de derechos humanos*, 13.

En cuanto a su estructura, la investigación prevé una primera sección dedicada a consideraciones generales sobre la igualdad y la no discriminación. A continuación, se desarrolla el tópico de la discriminación algorítmica en el ámbito laboral y se reflexiona en cómo el marco jurídico mexicano podría hacerle frente, para ello se identifican alternativas de lege data, así como retos y oportunidades de lege ferenda. Finalmente, se presentan consideraciones conclusivas.

II. CONSIDERACIONES GENERALES SOBRE LA IGUALDAD Y LA NO DISCRIMINACIÓN

Una de las grandes conquistas de la humanidad son los derechos humanos, entendidos como esferas irreductibles de protección para toda persona por el simple hecho de serlo y cuya justificación atiende al valor común de la dignidad inherente a la condición humana.¹⁷ A su vez, son el resultante de experiencias traumáticas a lo largo de la historia (como la Primera y la Segunda Guerra Mundial), así como una expresión de lo que desgraciadamente sigue ocurriendo en la realidad, pero que deseamos superar (hambre, desigualdad, guerras, violencia, etc.). Los derechos humanos son pues “mandatos de optimización” en los que resuena la voz de los desprotegidos, de quienes sufren carencias, violencia, exclusión y discriminación, en suma, son los “derechos del más débil” frente a la arbitrariedad y el poder de las fuerzas políticas y económicas de los más fuertes.¹⁸

Del mismo modo, y en esta visión, se asume que los Estados tienen como razón o justificación de su existencia la tutela de estos derechos que, inclusive, son reconocidos anteriores a éste y condicionantes a todo su actuar.¹⁹ Así, el Estado siempre estará obligado a tutelar a la persona

¹⁷ Ferrajoli, *Derechos y garantías. La ley del más débil*, 37; Nino, *Ética y derechos humanos. Un ensayo de fundamentación*, 43.

¹⁸ La referencia a “mandatos de optimización” corresponde a Alexy, *Teoría de los derechos fundamentales*, 67. La idea de los derechos humanos como leyes del más débil es de Ferrajoli, *Derechos y garantías...*, 54-55. Sobre los derechos humanos como la voz de los desprotegidos, véase: Williams, *La dolorosa prisión del lenguaje de los derechos*, 55.

¹⁹ Pinto, *Temas de derechos humanos*, 47-50.

humana como un fin en sí mismo y, por consiguiente, jamás podrá instrumentalizarlo.²⁰

Lo anterior es importante porque nos permite situar a los derechos humanos en su carácter de herramientas o instrumentos de tutela para quien se encuentra en situación desfavorable. Los derechos humanos plantean que, mediante diversas estrategias de adaptación, tales sectores de la población puedan acceder a un plano de auténtica igualdad y, como corolario, prohibir toda forma de discriminación.²¹

En el contexto contemporáneo subsisten las guerras, el hambre, la violencia y múltiples desigualdades, una de ellas se presenta en el ámbito tecnológico, concretamente en el de la IA y sus aplicaciones. Como hemos anticipado, ésta puede ser una herramienta que, en el caso de ser colocada para el beneficio de la humanidad, podría derivar en innumerables aspectos positivos; contrariamente, constituye un factor de riesgo susceptible de profundizar las desigualdades.²²

Sobre el efecto discriminatorio de la IA, concretamente en el ámbito de la discriminación laboral por algoritmos, se hablará más adelante, por lo pronto, conviene señalar que la discriminación puede presentarse de diferentes maneras. Para efectos de nuestro estudio nos interesan las categorías de discriminación directa e indirecta.

En su núcleo, toda forma de discriminación supone que, en condiciones análogas, sea brindado un trato diferencial. Cuando este trato se asienta en categorías como el color de piel, origen étnico, lengua, religión, opiniones, sexo o edad, hablamos de discriminación directa. Por la experiencia histórica, dichos criterios han sido considerados de antemano como contrarios al núcleo esencial de la dignidad humana, de modo tal que se erigen en categorías “sospechosas” que, prima facie, supondrían un trato discriminatorio.²³

Ahora bien, además de las categorías anteriores, existen aspectos o condiciones que aplicados en casos concretos pueden derivar en un trato

²⁰ Moccia, *Sistema penale e principi costituzionali: un binomio inscindibile per lo stato sociale di diritto. Relazione di sintesi*, 1725.

²¹ Habermas, *Facticidad y validez. Sobre el derecho y el Estado democrático de derecho en términos de teoría del discurso*, 499.

²² Abadías Selma, *El derecho penal frente a la discriminación laboral algorítmica*, 62-63.

²³ Capuzzo, “Do Algorithms dream about Electric Sheep? Percorsi di studio in tema di discriminazione e processi decisori algoritmici tra le due sponde dell’Atlantico”, 98.

diferencial, pero sin remitir a alguna categoría sospechosa. Aquí, ese criterio no necesariamente supondría un trato discriminatorio en todos los casos, sino únicamente atendiendo a la particular circunstancia que tiene lugar en cada situación. En este sentido se habla de discriminación indirecta o por impacto diferencial.²⁴ Lo que ocurre es que una categoría aparentemente neutra se aplica por igual a todos los casos sin tomar en consideración que determinados sujetos podrían resentir diversamente el impacto de la variable, sobre todo aquellos que se encuentran en una especial condición de vulnerabilidad. La discriminación (indirecta), en este caso es la consecuencia de no haberse valorado el impacto diferencial que esa medida podría haber supuesto para determinados sujetos.²⁵

En general, ante una discriminación directa el arsenal probatorio resulta sencillo de demostrar pues el criterio es objetivo, basta acreditar que a la persona se le ha tratado de manera diferente con base en alguna de las categorías sospechosas (que prima facie están prohibidas).²⁶ En la discriminación indirecta la cuestión es diversa porque se necesita demostrar el efecto diferencial que una categoría aparentemente neutra puede provocar en ciertos sujetos, además, el criterio no es objetivo pues la categoría per se es neutra, por ello será necesario acreditar un aspecto subjetivo consistente en la intención de discriminar (o al menos la omisión de un especial deber de prevención).²⁷ Como veremos, la presente clasificación será fundamental para el desarrollo de este estudio.

Otro aspecto que debe resaltarse desde ahora es que algunas formas de trato diferencial admiten justificación (por ello se habló antes del carácter prima facie). Como ya hemos adelantado, la filosofía de los derechos humanos reconoce que en el plano real existen desventajas estructurales para ciertos sectores de la sociedad. En consecuencia, se admite un trato diferencial que permita colocar en una condición de auténtica igualdad

²⁴ Tommasi, “Algoritmo e nuove forme di discriminazione: uno sguardo al diritto europeo”, 114-115.

²⁵ La categorización de discriminación directa e indirecta es un criterio asumido con rango de jurisprudencia por los tribunales mexicanos: Tesis 1a./J. 100/2017 (10a.), DISCRIMINACIÓN INDIRECTA O POR RESULTADOS. ELEMENTOS QUE LA CONFIGURAN. Gaceta del Semanario Judicial de la Federación, Décima Época, Libro 48, tomo I, noviembre de 2017, p. 225. Reg. Digital: 2015597, <https://sjf2.scjn.gob.mx/detalle/tesis/2015597>. Igualmente, Tesis 1a. CXXI/2018 (10a.), DISCRIMINACIÓN INDIRECTA O NO EXPLÍCITA. SU DETERMINACIÓN REQUIERE EL ANÁLISIS DE FACTORES CONTEXTUALES Y ESTRUCTURALES. Gaceta del Semanario Judicial de la Federación, Décima Época, Libro 58, tomo I, septiembre de 2018, p. 841. Reg. Digital: 2017989, <https://sjf2.scjn.gob.mx/detalle/tesis/2017989>

²⁶ Barbera, “Discriminazioni algoritmiche e forme di discriminazione”, 12-13.

²⁷ Gaudio, “Le discriminazioni algoritmiche”, 9.

real o material a quien no lo está. Este tratamiento es conocido como “discriminación positiva”²⁸ y el razonamiento que la justifica es que, si el ideario de la igualdad plantea un auténtico acceso y disfrute de los derechos humanos, pero en el plano fáctico subsisten obstáculos que imposibilitan acceder a esa auténtica igualdad, resulta entonces admisible la remoción de esos obstáculos, así como la incorporación de acciones idóneas, necesarias y proporcionales que permitan lograr una igualdad real. De otro modo, tratar por igual a quien es estructuralmente desigual no hace sino dar continuidad a la espiral de discriminación.²⁹ En estos casos, el tratamiento diferencial admite justificación, a condición de que se encuentre objetivamente justificado sobre la base de intereses legítimos y su aplicación sea idónea, necesaria y proporcional.³⁰

Otro aspecto de interés es que en muchos casos los derechos humanos podrán encontrarse en conflicto o colisión. Es decir, el ejercicio de un derecho puede al mismo tiempo suponer la limitación o restricción de otro. Cuando ambas categorías son derechos humanos resulta necesario realizar un “test de proporcionalidad” que nos indique en cada caso cuál de los derechos asume preferencia.³¹ Se trata de una estrategia interpretativa que posibilita el ejercicio pleno de los derechos cuando por determinadas circunstancias puedan verse en conflicto.

Los derechos humanos se asumen como la categoría de derechos de mayor importancia dentro de los ordenamientos jurídicos contemporáneos (democráticos y garantistas) de manera que, cuando en la hipotética colisión entre derechos uno tuviera el carácter de “humano” y el otro no, el primero tendría prevalencia. En el caso de nuestro estudio, se verá que son dables confrontaciones entre intereses como el respeto a la igualdad

28 Urteaga, “Las políticas de discriminación positiva”, 182-183.

29 Rivas Vallejo, *Discriminación algorítmica: detección, prevención y tutela*, 12.

30 Sáez Lara, “El algoritmo como protagonista de la relación laboral. Un análisis desde la perspectiva de la prohibición de discriminación”, 47-48. En el ámbito jurisprudencial de la Corte Interamericana de Derechos Humanos (Corte IDH) también se ha reconocido esta necesidad en diversos casos, entre otros: Corte IDH, Caso Guevara Díaz Vs. Costa Rica, “Fondo, Reparaciones y Costas. Sentencia de 22 de junio de 2022”, Serie C No. 453, 81, http://www.corteidh.or.cr/docs/casos/articulos/seriec_453_esp.pdf; Corte IDH, Caso V.R.P., V.P.C. y otros Vs. Nicaragua, “Excepciones Preliminares, Fondo, Reparaciones y Costas. Sentencia de 8 de marzo de 2018”, Serie C No. 350, 293, http://www.corteidh.or.cr/docs/casos/articulos/seriec_350_esp.pdf

31 Tesis 1a. CCLXIII/2016 (10a.), TEST DE PROPORCIONALIDAD. METODOLOGÍA PARA ANALIZAR MEDIDAS LEGISLATIVAS QUE INTERVENGAN CON UN DERECHO FUNDAMENTAL. *Gaceta del Semanario Judicial de la Federación*, Décima Época, Libro 36, tomo II, noviembre de 2016, p. 915. Reg. Digital: 2013156, <https://sjf2.scjn.gob.mx/detalle/tesis/2013156>. Asimismo: Alexy, *Teoría de los...*, 135-138; Alexy, *Tres escritos sobre derechos fundamentales y la teoría de los principios*, 73.

y la eficiencia económico-empresarial. A esta relación ha de sumarse un factor adicional y es el impacto que tecnologías como los algoritmos informáticos pueden tener para profundizar y maximizar la vulneración a los derechos humanos.³²

En nuestra opinión, la ecuación expresada habría de brindar supremacía a la tutela de la igualdad por lo que el desarrollo tecnológico sin frenos y su aplicación al ámbito laboral-empresarial únicamente podría permitirse en tanto fuere coherente con la tutela de los derechos fundamentales. Sin embargo, estas consideraciones parecerían no corresponderse en la realidad, pues asistimos a escenarios en donde el desarrollo tecnológico no se está traduciendo en un mundo que podamos considerar más justo y equitativo. En el caso concreto de la discriminación algorítmica veremos las sutiles pero poderosas herramientas con las que se excluye y se discrimina a amplios sectores de la sociedad.

III. LA DISCRIMINACIÓN ALGORÍTMICA EN EL ÁMBITO LABORAL

Una vez señalado el contexto general de nuestra investigación y los postulados básicos sobre la igualdad y la no discriminación, conviene presentar cómo es que la tecnología derivada de IA puede constituir un peligro para los derechos humanos.

La IA es un concepto difuso bajo el que se engloban una serie de técnicas informáticas que buscan emular o equiparar su funcionamiento (y resultados) a los de la inteligencia humana.³³ Dentro de la IA se distinguen diversas técnicas como el aprendizaje automático (Machine Learning),³⁴ el aprendizaje profundo (Deep Learning),³⁵ las redes bayesianas (Bayesian

³² Ramírez Aufrán, *Sesgos y discriminaciones...*, 9.

³³ Simón Castellano, *Justicia cautelar e inteligencia artificial. La alternativa a los atávicos heurísticos judiciales*, 48.

³⁴ El aprendizaje automático es la “rama de la Inteligencia Artificial que tiene como objetivo desarrollar técnicas que permitan a las computadoras aprender. De forma más concreta, se trata de crear algoritmos capaces de generalizar comportamientos y reconocer patrones a partir de una información suministrada en forma de ejemplos”, Rivas Vallejo, *Discriminación algorítmica...*, 3.

³⁵ “El aprendizaje profundo (*deep learning*) es un sistema de autoaprendizaje automático en el campo de las redes neuronales en inteligencia computacional, basado en el manejo de datos, que va construyendo capas de datos que alimentan y mejoran el algoritmo, lo que implica que el algoritmo aprende a partir de las distintas capas de datos que se le van introduciendo”, *Ibid.*, 7-8.

Networks)³⁶ o las redes semánticas (Semantic Networks).³⁷ Estas técnicas (y muchas más) requieren de un elemento básico que consiste en la secuencia o serie de instrucciones que les indican cómo deben operar, esa secuencia es el “algoritmo” y es aquí en donde reside fundamentalmente el aspecto susceptible de discriminación, por ello, para entender cómo es que puede darse un tratamiento de este tipo es importante conocer qué son y cómo funcionan los algoritmos.

1. ¿En qué consisten y cómo funcionan los algoritmos informáticos?

La Real Academia Española define al “algoritmo” como un “conjunto ordenado y finito de operaciones que permiten hallar la solución de un problema”.³⁸ En el ámbito informático vendría a ser ese conjunto de pasos, reglas o instrucciones que le indican a una computadora qué es lo que debe hacer para alcanzar un determinado objetivo.³⁹ Algunos autores resumen que el algoritmo es una suerte de receta de cocina o instructivo para ensamblar un juguete, sólo que, en este caso, aplicado a la informática.⁴⁰

Para que la computadora pueda alcanzar el logro trazado, esos pasos, reglas o instrucciones deben tener determinadas características, una de ellas es que sean lógicas y coherentes para evitar errores y contradicciones, así como para poder ser replicadas en cualquier otro ordenador con los mismos resultados.⁴¹ Otra regla es que las instrucciones deben ser finitas, es decir, limitadas en su extensión y secuencia, pues la idea de un algoritmo es que permita dar respuesta a un problema en el menor

³⁶ Estas redes “modelan un fenómeno mediante un conjunto de variables y las relaciones de dependencia entre ellas. Dado este modelo, se puede hacer inferencia bayesiana; es decir, estimar la probabilidad posterior de las variables no conocidas, en base a las variables conocidas. Estos modelos pueden tener diversas aplicaciones, para clasificación, predicción, diagnóstico, etc.”, Flores y Calderón, “Procesos de decisión de Markov y aprendizaje por refuerzo”, 376.

³⁷ “Una red semántica no es más que una notación gráfica para representar el conocimiento existente, diferenciando los conceptos y las relaciones y mostrando patrones de interconexión entre ellos [...] es una estructura de representación del conocimiento lingüístico”, Vilares Ferro *et al.*, “Ontologías y representación del conocimiento”, 105.

³⁸ “Algoritmo”, acceso el 15 de junio de 2024, <https://dle.rae.es/algoritmo>.

³⁹ Domingos, *The Master Algorithm*, 1.

⁴⁰ Merino Rus, *La igualdad en la era de los algoritmos y la inteligencia artificial*, 18; Mercader Uguina, “Algoritmos y derecho del trabajo”, 63.

⁴¹ Barbera, *Discriminazioni algoritmiche e...*, 7.

tiempo y con la menor cantidad de instrucciones posibles.⁴² Además, el algoritmo se encuentra “escrito” en un lenguaje específico que es de naturaleza matemática y que funciona con base en conectores de probabilidad.⁴³ Tal escritura no es fácil de comprender para un lego en informática, ni tampoco resulta accesible, pues en la mayoría de los casos esa información se encuentra protegida por el secreto empresarial (sobre este punto se hablará más adelante).

En la vida cotidiana gran parte de situaciones pueden expresarse como problemas de decisión y, por tanto, susceptibles de representarse mediante algoritmos. Si estas decisiones se repiten en el tiempo pueden ser automatizadas, de manera que un algoritmo podría emplearse para darle respuesta a situaciones como determinar cuál es la mejor ruta para llegar al trabajo, cuándo comprar un boleto de avión, dónde postularme como candidato a un puesto de trabajo o, desde el punto de vista del empleador, qué candidato es el que debería contratarse.⁴⁴

Ahora bien, otro aspecto de relevancia es el proceso mediante el cual funciona el algoritmo. Éste es el instructivo bajo el que se procesa, analiza, interpreta y se deriva una conclusión. Un algoritmo no equivale a la solución del problema, sino a la vía, mecanismo o procedimiento para llegar a su probable respuesta.⁴⁵ El algoritmo funciona con datos de entrada (inputs), que son sometidos a ese conjunto de reglas o instrucciones bajo las que el algoritmo fue programado y que terminan arrojando el resultado de salida (output).⁴⁶

El elemento central son los datos, sin ellos el algoritmo no es nada. A partir de la información *recibida*, el algoritmo deriva predicciones y esa información se asienta en “hechos”, en datos “duros” que han sido recopilados a lo largo del tiempo. La decisión algorítmica es la resultante del análisis histórico donde se nos presenta cuál ha sido la respuesta que

⁴² Merino Rus, *La igualdad en la era de los algoritmos...*, 18.

⁴³ Es importante precisar que el algoritmo no equivale al código fuente, aunque sí hace parte del mismo. El código fuente es el lenguaje bajo el que todo el programa informático se encuentra escrito y que define la interfaz, elementos gráficos, auditivos, componentes y reglas que “dan vida” al programa, incluyendo al algoritmo de trabajo, véase: Rivas Vallejo, *Discriminación algorítmica...*, 42; Domingos, *The Master Algorithm*, 2.

⁴⁴ Mercader Uguina, “Discriminación algorítmica y derecho granular: nuevos retos para la igualdad en la era del big data”, 5; Molina Navarrete, “Duelo al sol. ¿Un algoritmo controla mi trabajo? Sí; a tu empresa también”, 7.

⁴⁵ Tegmark, *Vida 3.0. Ser humano en...*, 97-105.

⁴⁶ Caterini, “El sistema penal en la encrucijada...”, 6.

se ha repetido el mayor número de veces frente a un problema específico. Esta evidencia crea patrones de comportamiento que luego pueden hacerse extensivos a situaciones iguales o similares.⁴⁷ Como se observa, el algoritmo no es magia ni nada que se le parezca, sino un análisis de estadística y probabilidad de donde se deriva su aludido potencial neutro o de “objetividad”.⁴⁸

En este procedimiento, mientras mayor sea la cantidad de información que le es suministrada, mayor probabilidad tendrá el algoritmo de que sus resultados se aproximen a la respuesta “correcta”. En este punto es donde las técnicas de IA contribuyen de manera sustancial a la optimización algorítmica.

En efecto, a través de la IA puede ponerse a disposición del algoritmo inmensas cantidades de información que han sido acumuladas a lo largo de los años, que se siguen generando y que por su volumen son imposibles de ser analizadas por la mente humana, se habla entonces del Big Data o de la macroinformación.⁴⁹ Además, la IA también posibilita optimizar el trabajo algorítmico a través de la incorporación de técnicas de Machine Learning y de Deep Learning (entre muchas otras). Con estas herramientas, aunque parezca difícil de creer, el algoritmo desarrolla la capacidad de aprendizaje (Machine Learning), que luego es permanentemente mejorada por el repetido filtrado de información en sus diversas capas de procesamiento (Deep Learning). El resultado de estos procesos permite al algoritmo optimizar su funcionamiento, lo cual supone modificar su programación originaria. Todo esto, además, de manera autónoma, es decir, sin la interferencia humana. Así, el algoritmo puede perfeccionarse a tal grado que, incluso, sea capaz de resolver el problema para el que ha sido creado en modos que ni siquiera hubieran sido anticipados por la mente humana.⁵⁰

⁴⁷ Rivas Vallejo, *Discriminación algorítmica...*, 6.

⁴⁸ *Ibid.*, 40.

⁴⁹ Mueller y Massaron, *Machine Learning for dummies*, 24-26.

⁵⁰ Un ejemplo clásico de la literatura se refiere al programa “AlphaGo”, diseñado con IA de Machine Learning para jugar al “Go” (un juego asiático similar al ajedrez de una antigüedad milenaria). En este caso, el programa comenzó aprendiendo de la información de programación y luego de la obtenida a partir de la experiencia. Al principio el programa era torpe, pero en cuestión de días mejoró sus estrategias hasta el punto en el que llegó a derrotar al mejor jugador en todo el mundo. En la partida final, el programa diseñó una estrategia que jamás había sido pensada y que tampoco estaba prevista en los datos de programación, lo cual nos conduce a la imprevisibilidad que un algoritmo de IA puede tener cuando llega a estos niveles de perfección, véase Tegmark, *Vida 3.0. Ser humano en...*, 97-105.

El potencial de las técnicas de IA aplicados a los algoritmos es inmenso. Con esta tecnología puede desarrollarse la capacidad predictora del algoritmo haciendo posible que trabaje con enormes bases de datos, analizando con escrupulosa minuciosidad la información existente y derivando patrones o asociaciones apoyadas en cientos, miles o incluso millones de casos semejantes ocurridos en el pasado.⁵¹

Dependiendo de la incorporación de tecnología de IA suelen distinguirse dos tipos de algoritmos: los de lógica deductiva (o deterministas) y los de probabilidad (o de Machine Learning). Los lógico-deductivos concentran sus esfuerzos en los datos de entrada de donde derivan reglas de generalización que luego aplican al caso particular (vía deductiva). En consecuencia, se trata de sistemas deterministas que: 1. Se apoyan en la lógica; 2. Son estáticos, porque el conjunto de operaciones es fijo y sólo puede ser modificado en fase de programación; 3. El resultado de salida es predecible a partir de los datos de entrada y; 4. El proceso decisional puede ser comprendido y explicado ex ante por la mente humana.⁵²

Los algoritmos de probabilidad funcionan de manera diferente, en este caso los esfuerzos no se concentran en los datos de entrada, sino en el conjunto de reglas u operaciones que tienen lugar en el algoritmo. Aquí, los inputs constituyen un primer acercamiento al problema que luego son complementados con Big Data y técnicas de IA. Este proceso permite derivar asociaciones, modelos y probabilidades que se aplican al caso concreto. De esta manera, la tecnología de aprendizaje automático permite que sea el propio algoritmo el que defina el conjunto de reglas u operaciones bajo las cuales serán tratados los datos de entrada para derivar los de salida. A su vez, la información de salida retroalimenta al algoritmo para que siga mejorando su capacidad de aprendizaje. Por ello, estos algoritmos tienen las siguientes características: 1. Se apoyan en probabilidades; 2. Son dinámicos, porque el conjunto de sus instrucciones se calibra y recalibra automáticamente con la experiencia (por tanto difieren de sus reglas originarias de programación); 3. Sus resultados no son predecibles porque derivan de una correlación de naturaleza probabilística que no siempre es causal y; 4. En la mayor parte de los casos, el proceso decisional es complejo y difícil de explicar, de hecho, técnicamente es

⁵¹ Esposito, "Note sparse sull'intelligenza artificiale", 37-50.

⁵² Gaudio, "Le discriminazioni algoritmiche", 5.

imposible realizar una valoración ex ante (porque no sabemos cuáles son las variables y el proceso decisional que el algoritmo tomó).⁵³

Como se advierte, en el caso de los algoritmos lógico-deductivos, el programador requiere escribir con detalle el conjunto de operaciones lógicas que el algoritmo deberá efectuar para poder derivar generalizaciones y luego las respectivas deducciones. En este proceso, sin embargo, se requiere conocer las relaciones causales que efectivamente produzcan el resultado deseado. En el caso de los algoritmos de aprendizaje automático no es necesario lo anterior, pues la propia máquina es capaz de identificar patrones, realizar asociaciones y luego cálculos de probabilidad. En este proceso, a la máquina sólo se le proporcionan algunos datos de programación que luego complementa y perfecciona con técnicas de IA de donde deriva sus propias predicciones (un proceder que, en efecto, se asemeja a la capacidad humana de aprendizaje guiada por la experiencia empírica).⁵⁴ Ahora bien, veamos a continuación cómo es que los algoritmos pueden ser empleados como “armas matemáticas” para la discriminación.⁵⁵

2. Potenciales vías de discriminación algorítmica

Como hemos dicho, la tecnología no es buena ni mala, sino funcional a los objetivos humanos, el caso de los algoritmos no es la excepción. Hemos visto ya qué son y cómo funcionan y de ahí podemos derivar ámbitos esenciales de cuya observancia dependerá que el algoritmo realmente proporcione la objetividad y fiabilidad que de éste se espera; caso contrario, el algoritmo puede ser un arma susceptible de afectar diversos derechos fundamentales, incluyendo la igualdad y su corolario de no discriminación. Tan preocupante es este aspecto que incluso se ha acuñado la expresión de “discriminación algorítmica” entendiéndose por ésta “la que se produce cuando un individuo o colectivo recibe un tratamiento arbitrario como consecuencia de la toma de decisiones automatizadas”.⁵⁶

⁵³ *Ibid.*, 6.

⁵⁴ Merino Rus, *La igualdad en la era de...*, 18.

⁵⁵ Para O’Neil, hoy en día muchos algoritmos son empleados como “armas de destrucción matemática” que se caracterizan por ser incontestables, secretos e injustos, O’Neil, *Weapons of Math Destruction...*, 3.

⁵⁶ Mercader Uguina, “Discriminación algorítmica y...”, 6. En sentido similar: Miró Llinares, “Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots”, 121.

Sobre este punto, podemos distinguir seis aspectos a los que deberemos prestar absoluta atención: el factor humano, la calidad de los datos, la representatividad, la utilización de proxys, la creación de perfiles o perfilación y el conocido efecto de “caja negra”.

2.1 El factor humano

Si bien es cierto que los algoritmos se expresan en reglas o secuencias matemáticas y que, por tanto, aspiran a una supuesta neutralidad, lo cierto es que todo algoritmo (de inferencia o de probabilidad) es programado, alimentado y entrenado por seres humanos. Este hecho jamás debe perderse de vista porque tanto el programador, el propietario o empresario que necesita al algoritmo dispone determinadas características que éste debe tener para ser funcional a sus intereses. De este modo, el algoritmo puede reflejar las opiniones, preferencias, gustos y demás caracteres subjetivos presentes en la condición humana que luego terminarán siendo aplicadas al proceso de análisis de datos.⁵⁷

Así, durante la fase de diseño o programación pueden ser incorporadas variables que eventualmente deriven en tratos diferenciales para cierta categoría de sujetos. Igualmente, estos sesgos pueden incorporarse en la fase de programación en donde se privilegie mejorar determinadas características a expensa de otras. Como ejemplo, en la mayor parte de los casos, los programadores son hombres, con una relativa posición económica favorable y con estudios en instituciones de determinado perfil (liberales, conservadoras, religiosas, laicas, etc.).⁵⁸ Todos estos factores incorporan un peso que de manera consciente o inconsciente se proyecta en las actividades del ser humano. Más específicamente podemos preguntarnos si durante su formación el programador recibió capacitación sobre la importancia de incluir la perspectiva de género o la necesidad de incorporar a grupos en condición de vulnerabilidad durante el análisis de los datos (personas de color, con preferencias sexuales diversas, pueblos

⁵⁷ Falletti, “Algoritmi: la discriminazione non è uguale per tutti”, 6.

⁵⁸ Gaudio, “Le discriminazioni algoritmiche”, 2.

indígenas, etc.),⁵⁹ o bien, si el programador experimentó algún evento traumático o situación compleja que pudiera condicionar su pensamiento y toma de decisiones. Estos elementos son fundamentales y deben ser identificados ex ante para tomar las debidas precauciones y evitar que esas subjetividades puedan ser trasladadas al algoritmo.⁶⁰

2.2 La calidad de los datos

El grado en el que la capacidad predictiva del algoritmo se aproxime a la realidad dependerá de la información con que éste haya sido alimentado. Por ello, resulta vital la calidad en la información y al logro de ésta contribuyen diversos elementos, entre otros: la cantidad con la que se ha alimentado al algoritmo (entre más casos con características similares, mayor será la probabilidad de conclusiones certeras); la representatividad de los datos (si sólo se le entrena a identificar cisnes negros, jamás identificará a un cisne blanco), la actualidad o vigencia de la información (si se le alimenta con datos donde las mujeres aún no se incorporaban ampliamente a la vida laboral, entonces no deducirá la importancia de la presencia femenina); la fiabilidad y completitud en la información (si la base de datos ha sido corrompida, entonces los resultados no serán objetivos).⁶¹

Además, debe realizarse una labor de “limpieza” en la información, de manera tal que nos aseguremos meticulosamente de que los datos incorporados son exactamente los que necesitamos y que no se han introducido de manera deliberada, o por accidente, categorías o elementos extraños que pudieran generar errores. En el ámbito informático, para referirse a este efecto suele emplearse la expresión *garbage in – garbage out*.⁶²

⁵⁹ Un caso conocido en donde no se incorporó la perspectiva de género y en donde además se evidenció una brecha salarial entre hombres y mujeres es el de Uber en Estados Unidos. En este caso, la compañía fue indagada en 2008 por el diferencial del 7% que sus trabajadoras ganaban en comparación con sus colegas hombres, asimismo, también fue indagada por diversas denuncias de acoso y hostigamiento sexual. En las indagatorias (realizadas por la Comisión para la Igualdad de Oportunidades) se acreditó que el algoritmo empleado no brindaba información sobre los criterios por los cuales en los hechos se producía ese diferencial del 7% (lo que, en el plano fáctico hacía que para el algoritmo el trabajo femenino fuera considerado con menor valor respecto del masculino), ni tampoco se demostró que el algoritmo hubiera sido entrenado en los riesgos diferenciales que el trabajo de autotransporte supone para hombres y mujeres. La compañía fue multada y se comprometió a modificar dicho algoritmo. Sobre este caso: Rivas Vallejo, *Discriminación algorítmica...*, 35-37.

⁶⁰ Barocas y Selbst, “Big data’s disparate impact”, 699-702.

⁶¹ Marcazzolo Awad y Walker Silva, “Empleo de sistemas algorítmicos de evaluación...”, 79.

⁶² Xenidis y Senden, “EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination”, 157.

Ciertamente, los algoritmos por sí mismos son incapaces de discriminar pues son reglas matemáticas que funcionan a partir de datos que le son suministrados. Sin embargo, cuando esos datos son inexactos, poco representativos, desactualizados o incongruentes, el resultado de su análisis será todo aquello. En este caso el algoritmo será falaz y poco confiable, además, si se le aplica a alguna decisión que pueda involucrar derechos fundamentales, el algoritmo pasará a ser una peligrosa arma para su lesión.

2.3 Representatividad

Este factor hace parte de la calidad de los datos, pero por su relevancia se le dedica una reflexión especial. En este caso debe cuidarse que durante el proceso de programación y entrenamiento al algoritmo le sea proporcionada información suficiente para que las correlaciones predictivas se ajusten a la realidad.⁶³ Así, por ejemplo, si a un algoritmo se le enseña a identificar patrones, pero únicamente se le entrena con una categoría de datos, será incapaz de detectar aquellos patrones con los que no se le entrenó.

Domingos llama a este factor el problema de la “ceguera” que eventualmente puede conducirnos a la “alucinación”. En el primer caso, el autor explica que un algoritmo sólo puede predecir aquello que ha conocido, es decir, si previamente se la ha suministrado esa información. En consecuencia, un algoritmo será “ciego” ante un evento o situación que no ha experimentado previamente, pero también podrá tener esa ceguera a pesar de que conozca ese evento o situación si los datos de configuración y entrenamiento respecto de esa variable son ínfimos o minúsculos. En este último caso, la información será insuficiente y parecerá que sólo existe una sola respuesta, dando lugar a una “alucinación”.⁶⁴

Un ejemplo de esta situación ocurrió en el caso de la compañía HireVue en los Estados Unidos la cual empleaba algoritmos de reconocimiento facial durante las entrevistas de trabajo. Sin embargo, en el proceso de entrenamiento algorítmico existió una insuficiente representación de mujeres y de personas de piel oscura; el resultado fue que, en la práctica,

⁶³ Miró Llinares, “Inteligencia artificial y justicia penal...”, 123.

⁶⁴ Domingos, *The Master Algorithm*, 70-75.

el programa resultaba “ciego” para poder detectar la cara y las facetas de mujeres no blancas.⁶⁵

Efectivamente, un aspecto que debe cuidarse en la programación y entrenamiento de los algoritmos es la suficiente representatividad, en especial de aquellos grupos que se encuentran en condiciones de vulnerabilidad o que históricamente han sido infra representados (pueblos indígenas, migrantes, personas con discapacidad, etc.) que, como podrá intuirse, muchas veces se encuentran fuera de las estadísticas o, en su defecto, los datos son apenas minúsculos en relación con otros colectivos.⁶⁶

Como hemos ya dicho, los algoritmos trabajan con estadística y si únicamente nos remitimos a ésta encontraremos que en la información de salida se privilegiará a determinadas categorías de personas en perjuicio de otras. Ello será así porque históricamente ha tenido lugar una jerarquización social de la que ha derivado una desigual distribución en los bienes, servicios y oportunidades (el fenómeno de la desigualdad material al que hemos aludido antes). Si al algoritmo se le deja únicamente con esa información sin realizarse los ajustes correspondientes (las acciones de discriminación positiva), a la postre sólo perpetuará modelos de discriminación que ya están presentes dentro de nuestras sociedades.⁶⁷

Por ejemplo, en los Estados Unidos, la estadística ha revelado que en las decisiones judiciales las personas de color suelen ser condenadas con mayor frecuencia frente a la población blanca. Esta presunción ha dado lugar a prejuicios que se continúan aplicando al impartir justicia y que también se perpetúan mediante programas algorítmicos que son alimentados con esas bases de datos. De esta manera, el programa aprende que esa pauta o patrón repetida a lo largo del tiempo es la opción más correcta y por ello concluye que será más probable el que una persona de color haya cometido un delito en comparación con una persona blanca.⁶⁸

65 Sáez Lara, “El algoritmo como protagonista...”, 44. Sobre este sesgo facial existe incluso un documental “Coded Bias” en donde se demuestra cómo ciertos sistemas de reconocimiento facial de grandes empresas, entre otras Amazon y Microsoft, funcionan de manera óptima con personas de rostro blanco y cómo, en el caso de personas de piel oscura, presentan diversas deficiencias, véase: Ferrante, “Inteligencia artificial y sesgos algorítmicos. ¿Por qué deberían importarnos?”, 28-29; Serna, Ignacio *et al.*, “Inside Bias: Measuring Bias in Deep Networks and Application to Face Gender Biometrics”, 3-5.

66 “La carencia de datos sobre sus condiciones particulares y singulares provocará que queden al margen, puesto que si el sistema algorítmico deja de contener patrones (datos) sobre los mismos, este los rechazará automáticamente y sin contemplaciones”, Asquerino Lamparero, “Algoritmos y discriminación...”, 376-381.

67 Miró Llinares, “Inteligencia artificial y justicia penal...”, 123.

68 Simón Castellano, *Justicia cautelar e inteligencia artificial...*, 119.

Es fácil inferir que el factor de la representatividad es fundamental en temas de derechos humanos porque la presencia histórica de estos derechos, de una parte, es una variable relativamente nueva en nuestras sociedades que, además, no se traduce efectivamente en una realidad (pues siguen existiendo innumerables vulneraciones a estos derechos). En correspondencia, si sólo empleamos información estadística en donde las decisiones han sido tomadas bajo estándares no armonizados con los derechos humanos, o bien, con datos objetivos pero que siguen revelando las falencias estructurales por las que se lesionan esos derechos, entonces las predicciones algorítmicas también arrojarán resultados que no se corresponderán con estos principios (pensemos en la perspectiva de género, el interés superior del menor, acciones afirmativas para determinados grupos, etc.). En tales casos, será necesario que el algoritmo “aprenda” esos conceptos por lo que deberá ser reprogramado y entrenado con información que muy probablemente se aleje del hecho factual (la injusticia y desigualdad del mundo exterior), pero que resultará necesario si queremos seguir tutelando los derechos humanos.⁶⁹

2.4 La utilización de proxys

Un proxy es una variable de aproximación que, en términos estadísticos y probabilísticos, encuentra correlación dentro de los datos de análisis.⁷⁰ La característica de esta variable es que no se determinan con base en una naturaleza causal, sino de probabilidad, por ejemplo, el hecho de que en una selección de personal se encuentre que la mayor parte de las personas elegidas viven en una determinada zona o que practican un determinado deporte, sin que estas variables hayan sido en modo alguno elementos predefinidos.⁷¹

Este tipo de correlaciones son propias de los algoritmos de Machine Learning y, a su vez, resultan especialmente peligrosas porque, eventualmente, pueden dar lugar a formas de discriminación, tanto directa como indirecta. En el primer caso, producto de una decisión al momento de la configuración algorítmica o, en el caso de la discri-

⁶⁹ Rivas Vallejo, *Discriminación algorítmica...*, 5.

⁷⁰ Asquerino Lamparero, “Algoritmos y discriminación...”, 366; Barocas y Selbst, “Big data’s disparate impact”, 720-721.

⁷¹ Tommasi, “Algoritmo e nuove forme di...”, 114-115.

minación indirecta, como resultante de un patrón que el algoritmo ha identificado por sí mismo.

Como ejemplo, se piense que el algoritmo “aprende” a identificar que, con base en la estadística criminal, las personas que en mayor grado son encontradas responsables viven en una determinada zona que puede delimitarse por el código postal. De esta manera, el código postal se identifica como una variable de proximidad, lo interesante es que por sí misma esta variable no nos dice nada y, en consecuencia, es neutra. Sin embargo, si realizamos una reflexión de mayor profundidad, podremos derivar el hecho factual de que las personas que viven en la zona identificada por ese código suelen ser personas de pocos recursos, incluso inmigrantes que prefieren habitar en esa zona porque los precios de los alquileres son más accesibles. Si ese proxy se aplica en la configuración del algoritmo lo que ocurrirá es que de manera indirecta (por derivación) se estará penalizando a la pobreza o incluso al origen nacional y, en estos casos, sí tendremos una variable que por sí misma resulta discriminatoria (la condición económica o la nacionalidad).

Como se observa en este caso, solamente tras un análisis posterior podemos darnos cuenta del factor discriminante en el proxy, cuestión que no siempre es sencilla de identificar. Por este motivo, la configuración de un proxy en los algoritmos resulta especialmente peligrosa y abre espacios para que de manera imperceptible puedan generarse consecuencias discriminatorias. Estas consecuencias pueden no ser deliberadas por parte del programador ya que sería el propio programa el que “aprenda” a identificar esos patrones. Sin embargo, pensando de una manera maliciosa, en otros casos sí que podría ser diseñado el algoritmo de esta manera, incorporando de manera velada alguno de estos proxys en la fase de programación o en la de entrenamiento. En cualquier caso, sea de manera deliberada o no, la tecnología incorporada supondrá múltiples dificultades para eventualmente acreditar una discriminación.⁷²

En estos casos, recuperamos el criterio de la discriminación indirecta que puede ser una importante herramienta para atajar el efecto adverso de medidas que, en apariencia, se presentan como neutrales pero que en los hechos proyectan efectos diferenciales para determinadas categorías de sujetos (normalmente en condiciones de vulnerabilidad). No obstante, se

⁷² Barbera, “Discriminazioni algoritmiche e forme...”, 12-13.

mantendrán algunas otras dificultades del ámbito probatorio, por ejemplo, justificar que la medida no ha sido idónea, necesaria o proporcional, que no se encuentra amparada bajo algún objetivo legítimo (por ejemplo, la libertad contractual del empleador o la eficiencia empresarial) y, además, que pueda eventualmente accederse a la información del algoritmo que permita explicar la raíz de la discriminación.⁷³

2.5 Creación de perfiles o perfilación

Los algoritmos son capaces de identificar patrones dentro de la información suministrada, con ellos elaboran generalizaciones y luego deducciones o cálculos de predicción, ¿qué sucede cuando dentro de estas generalizaciones es la persona humana la que se cuantifica?

En efecto, el algoritmo puede identificar patrones incluso dentro de la información que se refiere a nuestros datos personales. La estadística de estos datos permite generar un sistema de etiquetado y luego de creación de perfiles (profiling) sobre los que se derivan juicios a futuro.⁷⁴ Estos juicios predictivos se desprenden no de la singularidad del individuo, sino de su clasificación en una cierta categoría constituida por lo que otras personas con características como la raza, nacionalidad, posición económica, sexo u otras características han hecho en el pasado y de donde se derivan predicciones sobre lo que esa persona podría hacer en el futuro.⁷⁵

Además, no se pierda de vista que en todo este proceso el algoritmo efectúa un tratamiento automatizado de datos personales y que, en consecuencia, requiere (o requerirían) de la correspondiente autorización de su titular, así como la adopción de todas las medidas que se derivan de la protección de datos.⁷⁶

Como hemos adelantado, dentro de este sistema de perfilación pueden aplicarse presunciones sostenidas únicamente por estereotipos y también ser el resultado de un barrido general que no toma en cuenta la particular situación en la que se encuentra un sujeto o colectivo de cara al ordenamiento jurídico, en este último caso, las características

⁷³ Sáez Lara, “El algoritmo como protagonista...”, 48-49.

⁷⁴ Alameda Castillo, “Reclutamiento tecnológico. Sobre algoritmos y acceso al empleo”, 12-13.

⁷⁵ Miró Llinares, “Inteligencia artificial y justicia penal...”, 125.

⁷⁶ Tommasi, “Algoritmo e nuove forme di discriminazione...”, 118.

del sujeto pueden ser el resultado de su condición de vulnerabilidad, del cumplimiento de una determinada obligación o incluso del ejercicio de un derecho (por ejemplo, el no acudir a trabajar en el ejercicio del derecho constitucional a la huelga).⁷⁷

Como se observa, la perfilación supone el inconveniente de que no particulariza detalles específicos y en ello reside uno de sus potenciales efectos discriminatorios pues aplica consecuencias generales a sujetos que pueden encontrarse en una particular situación que obligaría a hacer excepciones.⁷⁸ Más aún, en el ámbito de los derechos humanos se plantean reflexiones ulteriores porque de la incorporación o no del sujeto en una determinada categoría derivan decisiones con incidencia en este tipo de derechos. Sin embargo, el parámetro de inclusión depende en lo absoluto de un sistema informático en el que no quedan del todo claras cuáles han sido las reglas y criterios que han supuesto la inclusión en esas categorías.⁷⁹

En este punto no podemos dejar de recordar que estrategias de este tipo han sido planteadas en el pasado pero descartadas por sus terribles efectos. En el ámbito criminal nos viene a la mente el recuerdo de la escuela positivista del derecho penal que en el siglo XIX proponía elaborar categorías de personas con base en sus rasgos físicos y psicológicos. De esta manera, podría apartarse de la sociedad a aquellas personas que desde su nacimiento tuvieran ciertos caracteres físicos o temperamentales (la idea del criminal nato de Cesare Lombroso).⁸⁰ Como vimos en el

77 Este fue el caso de la plataforma digital italiana Deliveroo que elaboraba perfiles de sus trabajadores con base en determinadas características donde se incluía la valoración de los usuarios y su disposición a laborar en determinadas horas pico. Sin embargo, la plataforma no distinguía criterios que permitieran a los trabajadores justificar su ausencia, incluso atendiendo a derechos constitucionales como el del derecho a huelga. La ausencia de esta distinción provocaba un trato diferenciado no justificado conforme a los parámetros constitucionales, razón por la cual fue declarada responsable de discriminación. Véase Tommasi, “Algoritmo e nuove forme di discriminazione...”, 116-117.

78 Barbera, “Discriminazioni algoritmiche e forme di discriminazione”, 10.

79 Este ha sido el criterio por el que algunos tribunales han determinado que la perfilación algorítmica no puede ser permitida. En España, el Ministerio para la Transición Ecológica empleaba un algoritmo para determinar el acceso (o exclusión) de la población a un bono eléctrico, sin embargo, muchas personas en situación de precariedad quedaban excluidas sin saberse muy bien por qué. El caso fue llevado a tribunales por la Fundación Civio y se concluyó que el algoritmo era una herramienta desproporcionada pues condicionaba la garantía de un derecho social a una entera decisión algorítmica (donde como consecuencia de la secrecía no quedaba en claro cuáles eran los criterios empleados por el algoritmo). Un caso similar ocurrió en Holanda donde a través de un sistema algorítmico se condicionaba el acceso a beneficios hacendarios. Para el juez, la creación de perfiles y la determinación de derechos con base en ello resultaba una medida carente de proporcionalidad y de transparencia. Sobre estos casos: Rivas Vallejo, *Discriminación algorítmica...*, 45-58; Morondo Taramundi, “Le sfide della discriminazione...”, 5.

80 Basile, “Intelligenza artificiale e diritto penale quattro possibili percorsi di indagine”, 10.

caso de los Estados Unidos y los programas algorítmicos de predicción criminal, en realidad esto ya ocurre, pero también en otros rubros. La consecuencia de todo ello es que, mediante la perfilación, los sistemas informáticos pueden arrojar consecuencias desastrosas para los derechos humanos, por ejemplo, predecir que una persona será criminal por su color de piel, que no podrá pagar un crédito por ser pobre o que tendrá un menor rendimiento laboral por el hecho de ser mujer.

Como se ve, la inclusión de la persona en una determinada categoría puede condicionar la decisión sobre el acceso a un trabajo, un crédito hipotecario, una ayuda económica o, incluso, la libertad personal. En el caso de los algoritmos lógico-deductivos los criterios de inclusión-exclusión pueden ocultarse hábilmente a través de una mayor complejidad del código algorítmico y, en el caso de los de Machine Learning la situación puede ser más compleja porque los criterios pueden ser desconocidos para el propio programador.

Otro factor a tener en cuenta es que en ninguno de estos procesos se respeta el libre arbitrio de la persona, ésta no tiene ninguna intervención, pasa a ser un mero instrumento del que se recopila información, tanto la que proporciona directamente como la que deja a lo largo de su huella en internet. Así, la persona es enteramente instrumentalizada vulnerando con ello su dignidad humana, pero también sus datos personales, su privacidad e incluso, el honor e imagen personal pues no se pierda de vista que el algoritmo crea un perfil, muchas veces estereotipado de la persona, del que deriva consecuencias.⁸¹

2.6 El efecto de “caja negra”

Antes hemos dicho que los algoritmos se encuentran escritos en un lenguaje particular y que, a su vez, hace parte del código fuente del programa. Pues bien, este código es en la mayor parte de los casos de acceso restringido porque se encuentra amparado por las normas en materia de propiedad industrial y por el secreto empresarial.⁸² La consecuencia es que, aunque podamos concluir que por sus efectos un determinado algo-

⁸¹ *Ibid.*, 15.

⁸² Morondo Taramundi, “Le sfide della discriminazione...”, 22.

ritmo puede ser discriminatorio, los procesos o reglas específicas de las que se derivó el trato diferenciado no las podemos demostrar.

En este sentido suele decirse que nos enfrentamos a una “caja negra” porque, aunque eventualmente podamos conocer cuál fue la información de entrada que se suministró al algoritmo y cuál fue su conclusión, no podemos acreditar el procedimiento específico bajo el que se llegó a esa decisión.⁸³

Un problema adicional es que, aún y cuando se pudiera acceder al código fuente, no podríamos entenderlo porque está escrito en un lenguaje matemático que sólo es descifrable para expertos en informática. De ahí que una de las propuestas que se han planteado sea la del reconocimiento de un derecho a la “explicabilidad” (en un lenguaje llano y comprensivo) del funcionamiento algorítmico.

Las complejidades de la caja negra no terminan aquí pues otro problema mucho más grave se presenta cuando nos enfrentamos a los algoritmos de Machine Learning. En este caso, recordemos que el algoritmo dispone de una autonomía que le permite constantemente autocorregirse para perfeccionar sus estrategias en la resolución de un problema. Apoyándose en esta característica, el algoritmo es capaz de derivar conexiones y patrones entre variables de gran complejidad para la mente humana, de manera tal que el algoritmo es sencillamente inexplicable para los seres humanos, pero, además, tampoco es posible conocerlo puesto que ya no es el algoritmo original de programación, sino uno que difiere en cada momento como consecuencia de su capacidad de autocorrección.⁸⁴ En este caso, el funcionamiento del algoritmo es todo un misterio.⁸⁵

⁸³ Pasquale, *The Black Box Society: The Secret Algorithms that control Money and Information*, 34.

⁸⁴ Esto fue lo que ocurrió precisamente a la compañía Amazon en 2014. La empresa diseñó un algoritmo para la contratación de su personal en Inglaterra, ese algoritmo fue entrenado con datos históricos de donde derivó una mayor propensión a contratar hombres frente a mujeres. El sesgo fue identificado y se procedió a reprogramar al algoritmo en diversas ocasiones; sin embargo, en cada ocasión continuaba arrojando sesgos de género. Al final, se optó por cancelar el algoritmo ante la imposibilidad de descifrar para los propios programadores el porqué se comportaba de esa manera el algoritmo (era un algoritmo de Machine Learning). Sobre este caso: Eguiluz Castañeira, “Desafíos y retos que plantean las decisiones automatizadas y los perfilados para los derechos fundamentales”, 345.

⁸⁵ “Cuanto más se perfecciona el algoritmo con la entrada de nuevos datos y el testeo de su mecanismo sobre ellos, más opaco se vuelve su mecanismo de funcionamiento interno. La transparencia de su red neuronal virtual se va diluyendo, escapando incluso al conocimiento de sus diseñadores, que lo entrenaron precisamente para que pudiera aumentar de manera automatizada su profundidad de análisis sin necesidad de intervención humana”, Rivas Vallejo, *Discriminación algorítmica...*, 9.

3. La discriminación algorítmica en el ámbito laboral. (Tratamiento automatizado) Elaboración

Lo dicho hasta aquí no debe conducirnos a una anticipada conclusión que pudiera llevarnos a condenar los procesos algorítmicos por su potencial discriminatorio. Se recuerda que la tecnología no es buena ni mala, sino funcional a las necesidades humanas. Por esta razón, también queremos señalar que esta tecnología puede aportar importantes beneficios al ámbito laboral-empresarial.

Uno de estos alicientes es la eficiencia. En efecto, dentro del ámbito empresarial una de las principales prioridades es la maximización de los beneficios a partir de una optimización de los recursos. El empleo de algoritmos informáticos permite desarrollar una suerte de análisis de gestión empresarial para determinar qué aspectos pueden optimizarse para obtener los mayores rendimientos. Así, por ejemplo, con base en los datos estadísticos pueden identificarse áreas de oportunidad, luego someterlas al análisis algorítmico que nos permita reducir pérdidas, aumentar los márgenes de utilidad, prevenir delitos y otros eventuales riesgos (laborales, ambientales, de imagen o prestigio, etc.).

Además, si se observa el debido cuidado en la calidad de los datos y en los demás criterios sobre los que antes hemos advertido, la consecuencia será que los sistemas algorítmicos permitirán a los decisores elegir aquellas alternativas en las que existe mayor evidencia empírica sobre sus resultados y beneficios. En consecuencia, podría alcanzarse esa aludida objetividad, certeza, precisión y ausencia de sesgos que en el ámbito laboral tampoco están exentos.

Sin embargo, de no observarse los criterios que antes hemos señalado, el escenario podría resultar muy diverso. En tales casos el algoritmo carecerá de fiabilidad y podría trasladar sus efectos negativos al ámbito de los derechos fundamentales, sobre todo de los trabajadores o trabajadoras en potencia (los candidatos en un proceso de selección). Pero, además, incluso respetando estos criterios, el empleador podría aplicar esta herramienta para evaluar la calidad de sus trabajadores en todo el proceso de la vida laboral (desde la contratación hasta la terminación de la relación de trabajo).⁸⁶ Así, mediante esta tecnología pueden valorarse

⁸⁶ Rivas Vallejo, *Discriminación algorítmica...*, 6.

aspectos como el rendimiento, la eficiencia o la calidad del trabajo. Más allá del empleo de dicha tecnología para estos fines (lo cual supone un necesario debate ético), lo preocupante es que si el algoritmo se encuentra sesgado, la decisión resultante incidiría en los derechos fundamentales con la agravante de que sería muy difícil de acreditar, colocando al sujeto en una condición de indefensión.

Atendiendo a las limitaciones en la extensión de esta investigación, así como al objetivo trazado (demostrar eventuales hipótesis de discriminación algorítmica en el ámbito laboral), concentraremos el estudio en uno de los momentos donde mayor probabilidad existe de ser objeto de discriminación algorítmica: el proceso de contratación.

Ramírez Bustamante y Páez comparten que, en 2018, el 98% de las compañías incluidas en el ranking Fortune 500 utilizaron sistemas algorítmicos para la contratación de personal,⁸⁷ también agregan que, en la actualidad, cerca del 50% de las medianas empresas europeas utilizan esta tecnología.⁸⁸ Además, los estudios encuentran coincidencia en que, con el paso del tiempo, los programas algorítmicos usados para la contratación de personal serán más accesibles permitiendo a toda empresa emplearlos como estrategia de productividad.⁸⁹

Estos sistemas algorítmicos se conocen como Applicant Tracking System o “ATS” y se aplican a lo largo de diferentes fases: la oferta laboral, la captación o reclutamiento de interesados, el envío y análisis de la información curricular, el desarrollo de entrevistas virtuales (que pueden suponer la interacción con algún Chatbot, el envío de un video o la captura de una imagen en tiempo real) y en la determinación de la decisión de contratación.⁹⁰

Primeramente, mediante algoritmos con tecnología de IA el programa es capaz de vincularse con el Big Data donde identifica los rastros que una persona ha dejado en la Web. A partir de ello, y sin que sea una casualidad, es como a determinada categoría de personas les “aparece” en su mail, redes sociales o en ventanas emergentes, una sugerencia o

⁸⁷ Ramírez Bustamante y Páez, “Análisis jurídico de la discriminación algorítmica en los procesos de selección laboral”, 2. Un porcentaje similar también es presentado en Rivas Vallejo, *Discriminación algorítmica...*, 28-29 y en Alameda Castillo, “Reclutamiento tecnológico...”, 18.

⁸⁸ *Ibid.*

⁸⁹ Bogen y Rieke, *Help wanted. An Examination of Hiring Algorithms, Equity and Bias*, 8-9.

⁹⁰ Rivas Vallejo, *Discriminación algorítmica...*, 26-27.

recomendación para postularse a una determinada oferta laboral (es el método que emplea Facebook, Twitter, InfoJobs y LinkedIn).⁹¹

Tras haberse posicionado para una audiencia personalizada, el sistema recibe los curriculum vitae de quienes han resultado atraídos por la oferta. Esta información nuevamente es analizada por los algoritmos que, mediante cribados iniciales, proceden a detectar palabras clave con las que se descarta o se da continuidad a los candidatos. Quienes “sobreviven” a esta primera selección algorítmica se les sujeta a análisis posteriores donde la información suministrada es complementada con la existente en redes sociales y con la “huella” o rastro digital dejado en el internet (al emplear motores de búsqueda, aplicaciones de geolocalización, imágenes o videos compartiros, comentarios y opiniones en redes sociales, etc.).⁹²

Como hemos adelantado, a veces el procedimiento puede contemplar la realización de entrevistas online, la interacción con asistentes virtuales (Chatbots) o el envío de imágenes y video en tiempo real. El empleo de estas herramientas tampoco es casual, puesto que a través de nuestra imagen y movimientos corporales los algoritmos pueden predecir aspectos como la inteligencia emocional, aptitudes sociales, ideología, personalidad y otras habilidades blandas.⁹³ También puede predecirse si la persona dice o no la verdad.⁹⁴

A partir de ello el programa elabora categorías o perfiles de los candidatos estableciendo a su vez “rankings” con base en las cualidades o parámetros predefinidos para cubrir la vacante.⁹⁵ Como podrá advertirse, a lo largo de diversos momentos pueden presentarse escenarios en los que, de no estar adecuadamente programado y entrenado el algoritmo, pueden presentarse escenarios para la discriminación.

Un primer momento corresponde a la definición de los criterios o atributos bajo los cuales se define el prototipo de empleado a contratar. En este rubro, el empleador dispone de una libertad contractual para

⁹¹ Alameda Castillo, “Reclutamiento tecnológico...”, 18; Barocas y Selbst, “Big data’s disparate impact”, 683.

⁹² Bogen y Rieke, , *Help wanted. An Examination...*, 40-41.

⁹³ Asquerino Lamparero, “Algoritmos, procesos de selección y reputación digital: una mirada antidiscriminatoria”, 128.

⁹⁴ Fita Ortega, “El derecho a la mentira como tutela preventiva de los derechos fundamentales de los trabajadores”, 97-99; Rivas Vallejo, *Discriminación algorítmica...*, 26-27.

⁹⁵ Asquerino Lamparero, “Algoritmos, procesos de...”, 122.

definir a quién incorpora y a quién no, pero, para que su decisión sea coherente con el ordenamiento jurídico, y en particular con los derechos humanos, su decisión no debe ser arbitraria ni discriminatoria. Así, por ejemplo, si dentro de los requisitos se dispone alguna categoría sospechosa que no pueda justificarse de manera objetiva, por ejemplo, contratar únicamente hombres, persona de una cierta franja de edad o de un cierto color de piel, la decisión será discriminatoria (hipótesis de discriminación directa).

En la búsqueda del empleado modelo, el empresario puede acudir a diversas estrategias. Una de ellas es el análisis histórico de la estadística. En este caso, con el objetivo de lograr la máxima eficiencia, podría pensarse en aquellos trabajadores históricos que mejor desempeño han tenido, una vez identificados se derivan sus caracteres y se incorporan como modelos en la programación algorítmica. En sentido inverso, también puede identificarse a los empleados con peor desempeño, tanto aquellos actuales como aquellos pasados, derivando luego los caracteres que no se desean contratar. Esta técnica permite un sistema de etiquetado por el que el algoritmo identificará factores de valoración positiva y de descarte.⁹⁶

Ahora bien, la discriminación puede presentarse por razones como la baja representatividad en los datos, la antigüedad de éstos, la detección de un carácter que pueda constituir una categoría sospechosa o que pudiera remitirnos a una de ellas por vía indirecta (proxy), así como por la ausencia de una programación específica que contemple, por ejemplo, cuotas para grupos en situación de vulnerabilidad.

Otra situación problemática tiene que ver con los criterios por lo que los sistemas de ATS identifican sectores de la sociedad a los que podrían estar privilegiando. Hemos ya dicho que no es ninguna coincidencia el hecho de que, por ejemplo, a ciertas personas se les ofrezca determinadas ofertas de trabajo y a otras no. Esto tiene que ver con nuestra información disponible en diversas bases de datos a la que el empleador puede tener acceso y que, tras un primer cribado, le presentan candidatos compatibles con los criterios de interés empresarial. El problema reside en que el criterio de exclusión no es transparente y en muchos casos se encuentra

⁹⁶ Kleinberg *et al.*, “Discrimination in the Age of Algorithms”, 132-136.

guiado por sesgos y prejuicios. Dentro de estas ofertas, por ejemplo, no es raro que sean excluidas personas migrantes, de color o mujeres.⁹⁷

En un estudio efectuado en 2015 se demostró que el programa de publicidad de Google, “Google Ads”, mostraba con una mayor frecuencia los empleos con mayor remuneración y estatus social a hombres que a mujeres (un promedio de 6 a 1), en el caso concreto, cargos ejecutivos en importantes empresas.⁹⁸ En otro estudio similar, se corroboró que Facebook hacía lo mismo, en este caso, del total de los destinatarios, las ofertas vinculadas con actividades empresariales de gran remuneración se mostraron 72% a hombres blancos; ofertas para trabajar como cajera en un supermercado 85% a mujeres y; ofertas para la conducción de taxis 75% a hombres afroamericanos.⁹⁹

Como hemos adelantado, los proxys son también una vía para la discriminación. Como se recordará, el algoritmo es capaz de identificar un patrón que, en principio, parecería neutro pero que a la postre podría reconducirse a alguna categoría sospechosa, de manera que, al final, se terminaría generando una situación de trato diferencial porque ciertos sujetos resienten de manera diferente ese patrón con base en las condiciones históricas y estructurales de desigualdad en las que se han encontrado. De nueva cuenta, pensemos en el código postal, en este caso el algoritmo podría identificar como datos históricos del “mal empleado” a aquellas personas que llegan tarde y que, además, viven en una determinada zona. Puede ocurrir que esa zona sea habitada mayoritariamente por personas de escasos recursos y, que, como suele ocurrir, esté apartada de los grandes centros de negocios. La explicación de segundo grado que debemos hacer (y que el algoritmo no nos dice) es que una persona con dificultad económica elige esa zona porque no tiene otra opción y que, en consecuencia, podría tener algunos problemas de puntualidad porque vive en una zona lejana, con problemas de comunicación con el centro urbano, porque se inunda constantemente, porque con frecuencia hay

⁹⁷ “Si nos situamos desde el lado de la plataforma o de las aplicaciones y relacionamos a éstas con el empleo de algoritmos, se observa cómo ellos son el instrumento perfecto para extraer las preferencias laborales de los candidatos y, en función de estas opciones laborales, la plataforma está capacitada para dirigirse a cada persona ofreciéndole las vacantes,” Asquerino Lamparero, “Algoritmos, procesos de...”, 129.

⁹⁸ Datta *et al.*, “Automated Experiments on Ad privacy Settings: A Tale of Opacity, Choice, and Discrimination”, 13-16.

⁹⁹ Ali *et al.*, “Discrimination through optimization: How Facebook’s ad delivery can lead to skewed outcomes”, 2-13.

episodios de violencia generalizada, etc. De no realizarse esta valoración e incorporar de manera directa los proxys, el algoritmo terminará por valorar negativamente un aspecto que puede ser reconducido a la variable de la condición económica y, en consecuencia, podría plantearse una hipótesis de discriminación indirecta. Además del código postal o el grado de puntualidad, existen otros proxys aparentemente neutros, pero con potenciales efectos discriminatorios como el apellido o el nombre de pila (de donde puede inferirse por probabilidad la nacionalidad, religión u origen étnico).¹⁰⁰

Reiteramos que debe tenerse un enorme cuidado en el empleo de proxys porque pueden ser la vía para ocultar formas de discriminación. Una empresa bienintencionada, por ejemplo, podría disponer que en los currículos de todos los candidatos se evitarán referencias al sexo, edad, nacionalidad u origen étnico, es decir, una suerte de “currículo ciego” en aras de evitar toda forma de sesgos.¹⁰¹ Sin embargo, de poco serviría ocultar esos datos cuando, mediante la tecnología de IA, el algoritmo podría derivarlos y aplicarlos a partir de una serie de conexiones, asociaciones e inferencias (donde han estudiado los candidatos, cuál es su código postal, palabras empleadas en el currículo, hobbies y la huella dejada en internet).¹⁰²

En el caso contrario, puede pensarse en un malintencionado empleador que en su fuero interno deseara contratar a personas que pertenezcan a una cierta edad, sexo y nacionalidad, pero claramente no deseara ser sancionado por este tipo de decisiones. En este caso, puede requerir al programador que entrene al programa o que introduzca de manera sumamente velada alguna variable aparentemente neutral que, sin embargo, privilegie a esa categoría de candidatos que desea contratar. En este panorama, el empleador dispone de muchas posibilidades para proceder como desea y no verse afectado pues, podría decir que la decisión se toma con base en un objetivo legítimo que es la mayor eficiencia empresarial y

100 En otro estudio, se demostró que el buscador de anuncios de Google fue capaz de predecir con exactitud que un 88% de los candidatos a un puesto de trabajo eran de piel oscura, mientras que un 96% de piel blanca. El factor empleado para predecir lo anterior fue tan solo el nombre de pila del candidato, véase Sweeney, “Discrimination in Online Ad Delivery”, 10-29. Igualmente, Kosinski, Stillwell y Graepel, “Private traits and attributes are predictable from digital records of human behavior”, 5802.

101 Fernández García, “La lucha contra la discriminación en el acceso al empleo mediante el curriculum vitae anónimo”, 244-249.

102 Asquerino Lamparero, “Algoritmos, procesos de...”, 126.

que además se trabaja con categorías neutras apoyadas en la estadística. En este supuesto, la persona o personas afectadas difícilmente podrían contrariar estos argumentos porque, aunque se sospeche que ha existido una discriminación (en algunos casos la exclusión puede ser tan sutil que pase por inadvertida),¹⁰³ ésta no se puede demostrar porque no se tiene acceso ni a los datos de entramiento, ni al código del algoritmo por vía del secreto empresarial.

Además, como ya hemos visto, aunque remotamente se tuviera acceso al código fuente, esta información casi siempre será incomprensible para el afectado por lo que requerirá un peritaje específico al respecto. En algunos otros casos, podrá ser incluso imposible detectar cuál ha sido el factor o factores que en lo específico han sido privilegiados por el algoritmo, así como el razonamiento o correlaciones que lo han llevado a tomar esa decisión porque puede tratarse un algoritmo de Machine Learning donde nadie, incluso el empleador o programador, podrían saber bien a bien cómo es que el programa funciona. En este último caso se presume que la conclusión es la mejor opción identificada por el algoritmo, pero no sabemos en realidad por qué y es esa ausencia de comprensión o explicación lo que convierte en arbitraria a la decisión (el efecto caja negra).¹⁰⁴

Un factor adicional de peligro para los derechos humanos que en cierto sentido ya hemos avanzado es que la tecnología de IA permite el tratamiento de bases de datos masivas de donde obtiene información sobre nosotros, un proceder que al menos plantea dudas en términos de privacidad y tutela de los datos.¹⁰⁵ En todo este proceso, no sabemos si la información se obtiene de bases públicas o privadas, ni tampoco qué se hace con nuestra información, dónde se guarda, quién la custodia, si previamente se nos requirió por alguna vía o no. Desconocemos si, además, el algoritmo ha tenido acceso a información sumamente privada como nuestros historiales médicos, estado civil, hábitos saludables o tóxicos, preferencias sexuales o círculo de amigos.¹⁰⁶

103 Morondo Taramundi, “Le sfide della discriminazione...”, 28.

104 Núñez Zorrilla, *Inteligencia artificial y responsabilidad civil. Régimen jurídico de los daños causados por robots autónomos con inteligencia artificial*, 12.

105 Sobre este punto en el ámbito de sectores sociales especialmente vulnerables: Abadías Selma, *Child Grooming: el embaucamiento de menores en la era del Metaverso y la inteligencia artificial*, 14.

106 Olarte Encabo, “La aplicación de la inteligencia artificial a los procesos de selección de personal y ofertas de empleo: impacto sobre el derecho a la no discriminación”, 83-85.

Sobre esta información también caben cuestionamientos en relación con su idoneidad para una cuestión que debería ser estrictamente laboral. Ciertamente, el empleador puede valorar aspectos como el nivel de estudios o las técnicas y habilidades que acrediten el desempeño para el trabajo en cuestión, algo completamente diferente es que, con nuestros datos personales el algoritmo pueda predecir cuestiones que, en primer lugar, no están comprobados que sucederán (son predicciones) y, en segundo lugar, que no hacen parte del ámbito estrictamente laboral.¹⁰⁷ Así, por ejemplo, el algoritmo puede realizar valoraciones sobre nuestro temperamento, tendencias a la depresión, grado de “felicidad”,¹⁰⁸ confiabilidad o lealtad, miedos e inseguridades, permanencia a largo tiempo en la empresa, posibilidad de tener o de desarrollar alguna enfermedad, de faltar con cierta frecuencia o, inclusive, de poder quedar embarazada (por ejemplo, a través de las búsquedas que sobre estos temas la candidata pueda hacer en internet).¹⁰⁹

A la par, no perdamos de vista los problemas de la perfilación en el ámbito laboral. En este caso, puede ubicarse a un candidato en una determinada categoría de sujetos, por ejemplo, los que podrían llegar tarde, enfermarse, deprimirse o resultar conflictivos. En tales supuestos, la predicción se hace a partir de datos estadísticos que, como hemos visto, pueden partir de prejuicios. El problema de la perfilación no termina aquí porque cuando un candidato ha sido objeto de un tratamiento automatizado que le arroja una valoración negativa, esa información no desaparece, sino que pasa a formar parte de un dato adicional con el que el programa se sigue entrenando. Sin embargo, ese candidato dispone ahora de un historial negativo contenido en una base de datos en internet. Esta información, de hecho, muchas veces es compartida con otros sistemas informáticos de selección laboral, de manera que, con ese score o valoración negativa la persona queda “fichada” y toda valoración posterior que pueda hacerse de ella bajo un sistema de datos que esté interconectado le terminará por arrojar sólo resultados negativos.¹¹⁰

107 De Vicente Pachés, “Acoso y discriminación en el trabajo a través de dispositivos tecnológicos o digitales: su incidencia en la fase de selección y acceso al empleo”, 217-219.

108 Es el caso del algoritmo “ADNe” que mediante diversas técnicas puede derivar el grado de felicidad de un trabajador, véase: Rivas Vallejo, *Discriminación algorítmica...*, 33.

109 Fita Ortega, “El derecho a la mentira como tutela...”, 99.

110 Ramírez Bustamante y Páez, “Análisis jurídico de la discriminación...”, 21.

En suma, como puede advertirse, el empleo de los algoritmos aplicado a la fase de contratación laboral plantea muchos aspectos en donde conviene prestar la debida atención. En este caso, la libertad empresarial para la contratación puede encontrarse en tensión frente a derechos fundamentales como la igualdad y la no discriminación, pero también dar margen a la lesión de derechos adicionales como la privacidad, intimidad, la tutela de los datos personales, la tutela judicial efectiva (se recuerde el efecto caja negra), la propia imagen e, incluso, el honor.¹¹¹ Llegados a este punto pasemos a analizar de qué recursos puede valerse nuestro ordenamiento jurídico para hacer frente a los efectos que la tecnología algorítmica puede tener en los derechos humanos, así como alternativas o posibilidades que a futuro pudieran considerarse para hacer frente a esta situación.

IV. MARCO NORMATIVO CONTRA LA DISCRIMINACIÓN LABORAL ALGORÍTMICA

La prohibición de la discriminación es un principio basilar de nuestro ordenamiento jurídico a tal grado que se reconoce desde el artículo 1.º, párrafo quinto, de la CPEUM al establecer la prohibición de “toda discriminación” que se encuentre motivada por:

[...] origen étnico o nacional, el género, la edad, las discapacidades, la condición social, las condiciones de salud, la religión, las opiniones, las preferencias sexuales, el estado civil o cualquier otra que atente contra la dignidad humana y tenga por objeto anular o menoscabar los derechos y libertades de las personas.¹¹²

La disposición en cuestión incorpora las “categorías sospechosas” a las que ya hemos hecho referencia; sin embargo, lo que prohíbe es una discriminación “motivada” por alguna de esas razones, es decir, debe demostrarse o acreditarse que ha existido ese motivo, cuestión

¹¹¹ El honor puede verse afectado cuando el algoritmo crea una imagen o percepción de una persona (un perfil) que podría apoyarse en estereotipos denigrantes para el individuo (holgazán, faltista, inadaptado, etc.), véase: Núñez Zorrilla, *Inteligencia artificial y responsabilidad civil.*, 11.

¹¹² Art. 1.º, párrafo quinto, Constitución Política de los Estados Unidos Mexicanos, disponible en <https://www.diputados.gob.mx/LeyesBiblio/pdf/CPEUM.pdf> (acceso el 20 de junio de 2024).

que como hemos ya adelantado resulta compleja en el caso de la discriminación algorítmica.

Igualmente, el artículo 123, Apartado A, fracción XXVII, inciso H, de la Constitución, dispone como una condición nula en la relación de trabajo toda estipulación que pueda implicar una renuncia de derechos consagrados en favor del obrero “en las leyes de protección y auxilio de los trabajadores”. En este caso, por ejemplo, podría pensarse en los ya aludidos derechos al honor, la imagen personal, la privacidad, la protección de datos y la prohibición de discriminación. Esta disposición de principio podría resultar aplicable a las hipótesis que previamente hemos señalado, sin embargo, si nos atenemos al tenor literal, la norma dispone esa cláusula para la categoría de un trabajador y, en los procesos de selección laboral, nos encontramos frente a personas que aún no reúnen esa calidad.

En el ámbito de la legislación secundaria, la Ley Federal del Trabajo (LFT) también incorpora normas de principios aplicables. En su artículo 2.o, párrafo primero, expresa que las normas laborales “tienden a conseguir el equilibrio entre los factores de la producción y la justicia social”, así como a “propiciar el trabajo digno o decente en todas las relaciones laborales”. De particular relevancia es el párrafo segundo donde se define al trabajo digno como:

[...] aquél en el que se respeta plenamente la dignidad humana del trabajador; no existe discriminación por origen étnico o nacional, género, edad, discapacidad, condición social, condiciones de salud, religión, condición migratoria, opiniones, preferencias sexuales o estado civil [...].¹¹³

Dicha disposición podría ser empleada para hipótesis de discriminación directa en tanto la preposición “por” permite inferir que el hecho discriminatorio deriva de alguna de las categorías sospechosas que el texto prevé, pero, en consecuencia, plantearía dificultades para la hipótesis de discriminación indirecta.

El mismo artículo 2.o también dispone que “Se tutela la igualdad sustantiva o de hecho de trabajadores y trabajadoras frente al patrón”.

¹¹³ Art. 2.o, párrafo segundo, Ley Federal del Trabajo, disponible en <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFT.pdf> (acceso el 20 de junio de 2024).

Esta previsión, en correlación con los párrafos primero y segundo del mismo artículo, nos llevaría a la conclusión de que el empleo de instrumentos tecnológicos como los algoritmos sólo podrían ser usados tras un análisis de ponderación en donde efectivamente se acredite que existe un equilibrio tanto en los intereses empresariales como en los derechos de los trabajadores, en especial su igualdad real o material. No perdamos de vista que, de cualquier manera, nos seguimos refiriendo con exclusividad a “trabajadores”.

Artículos adicionales de la LFT que prohíben los tratos discriminatorios son: 3, segundo párrafo; 56; 133, fracción I; 685 Ter, fracción I y; el 857, fracción IV. Sin embargo, en todos estos casos también se incorpora el requisito de “motivar” o acreditar que la discriminación sea “por razón” de alguna categoría sospechosa, cuestión que, reiteramos, resulta difícil de demostrar con la tecnología algorítmica.

Igualmente, es de resaltar que ni la CPEUM ni la LFT hacen mención alguna a conceptos como los de inteligencia artificial o algoritmos,¹¹⁴ aunque como dato curioso la LFT hace referencia a precauciones especiales frente a “cualquier tecnología” en el artículo 330-I donde indica:

Los mecanismos, sistemas operativos y cualquier tecnología utilizada para supervisar el teletrabajo deberán ser proporcionales a su objetivo, garantizando el derecho a la intimidad de las personas trabajadoras bajo la modalidad de teletrabajo, y respetando el marco jurídico aplicable en materia de protección de datos personales.¹¹⁵

En términos de nuestra investigación, la norma es relevante, no obstante, sus efectos se limitan a la relación laboral de teletrabajo, cuestión que es diversa a lo que hemos reseñado con antelación.

La Ley Federal para Prevenir y Eliminar la Discriminación (LFPED) dispone en su artículo 1.º la prohibición de la discriminación como un principio fundamental y, a la par (como lo hemos hecho en esta investigación), vincula esa prohibición a la necesidad de garantizar una auténtica

¹¹⁴ Tampoco la Ley Federal de los Trabajadores al Servicio del Estado, reglamentaria del Apartado B) del artículo 123 Constitucional. Si bien es cierto nos hemos concentrado en la potencialidad discriminatoria de los algoritmos en el ámbito privado, todo lo señalado también puede aplicarse al ámbito de la contratación pública.

¹¹⁵ Art. 330-I, Ley Federal del Trabajo.

igualdad real o material.¹¹⁶ En ese mismo artículo, en su fracción III, se incluye una definición amplia de discriminación:

[...] toda distinción, exclusión, restricción o preferencia que, por acción u omisión, con intención o sin ella, no sea objetiva, racional ni proporcional y tenga por objeto o resultado obstaculizar, restringir, impedir, menoscabar o anular el reconocimiento, goce o ejercicio de los derechos humanos y libertades, cuando se base en uno o más de los siguientes motivos: el origen étnico o nacional, el color de piel, la cultura, el sexo, el género, la edad, las discapacidades, la condición social, económica, de salud física o mental, jurídica, la religión, la apariencia física, las características genéticas, la situación migratoria, el embarazo, la lengua, las opiniones, las preferencias sexuales, la identidad o filiación política, el estado civil, la situación familiar, las responsabilidades familiares, el idioma, los antecedentes penales o cualquier otro motivo.¹¹⁷

Esta definición es ideal para hacer frente a las diversas hipótesis de discriminación algorítmica. En efecto, la definición reconoce que un tratamiento discriminatorio puede ser el resultado de motivar una conducta, por acción u omisión, en alguna de las categorías sospechosas, pero, además, reconoce que el tratamiento diferencial se puede demostrar por sus efectos o resultados, y esta situación es fundamental porque aquí lo que debe acreditarse no es el motivo en el que se apoya el trato discriminatorio, sino la consecuencia, efecto o resultado que produce. Además, la norma es relevante porque reconoce formas de comisión activas y omisivas, con intención o sin intención y aplicables a toda persona (no sólo a trabajadores). Se agrega a lo anterior el reconocimiento explícito que se hace en la fracción II Bis, incisos a) y b), de la discriminación directa e indirecta.

En la misma Ley, también debe señalarse el artículo 9, fracción III, que se refiere específicamente al ámbito laboral entendiendo que será un trato discriminatorio, el “[p]rohibir la libre elección de empleo, o restringir las oportunidades de acceso, permanencia y ascenso en el mismo”. Como hemos visto, la tecnología algorítmica aplicada a este rubro podría ser una hipótesis en la que se verían restringidas las oportunidades de acceso al empleo.

¹¹⁶ A tal efecto la LFPED incorpora “medidas de nivelación” (artículo 15 Ter), “medidas de inclusión” (artículo 15 Quáter) y “acciones afirmativas” (artículo 15 Séptimus).

¹¹⁷ Art. 1, fracción III, Ley Federal para Prevenir y Eliminar la Discriminación, disponible en <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPED.pdf> (acceso el 20 de junio de 2024).

Ahora bien, la LFPED tampoco hace referencia a la IA o a los algoritmos, aunque observamos que una discriminación algorítmica podría englobarse dentro del concepto genérico de discriminación que aquí se incorpora. Sin embargo, subsistirían obstáculos tanto generales como específicos. En primer término, frente a cualquier tipo de discriminación algorítmica, la LFPED es incapaz de ampliar su ámbito de atribuciones a aspectos que no sean de índole federal y tampoco puede derivar tipos de responsabilidad diferentes a los de naturaleza administrativa (considérese, por ejemplo, en la discriminación algorítmica derivada de un programa informático de predicción criminal). De otra parte, en el ámbito específico de la discriminación laboral algorítmica, la LFPED tampoco nos brinda reglas respecto de cómo identificar responsabilidades por este tipo de discriminación, es decir, aunque pudiera subsumirse la discriminación algorítmica en el concepto genérico de la discriminación, no sabemos quién es la persona que debe responder (el empleador, propietario o programador) o, incluso, si cabría alguna forma de responsabilidad para un sujeto no humano (se recuerden los algoritmos de Machine Learning con autonomía decisional).

En el ámbito estrictamente científico y tecnológico, nuestra legislación federal dispone de la Ley General en Materia de Humanidades, Ciencias, Tecnologías e Innovación (LGMHCTI), la cual fue publicada en el Diario Oficial de la Federación (DOF) apenas el 8 de mayo de 2023. En ella también se reitera la prohibición de toda forma de discriminación para hacer efectivo el derecho humano a la ciencia (artículo 1.º, párrafo tercero), sin embargo, también se acude al criterio de la discriminación que debe ser demostrada por haber estado motivada en alguna categoría sospechosa. Lo relevante es que una Ley tan actual y referida al ámbito científico omita toda referencia a la IA y a los algoritmos.

Por vía de la protección de datos nos encontramos con herramientas que son de una gran utilidad. La Ley General de Protección de Datos en Posesión de Sujetos Obligados (LGPDPDO) reconoce los derechos de acceso, rectificación, cancelación y oposición (artículos 12, fracción II y 43 a 56) que sin duda alguna son valiosos instrumentos para hacer frente a la discriminación algorítmica. La ley, además, incorpora otra poderosa herramienta en su artículo 47, fracción II, donde se hace referencia a los sistemas automatizados que, como hemos visto, incluyen a los algoritmos. En este caso se refiere la existencia de un derecho

del titular de estos datos para oponerse a su tratamiento o para exigir que éste cese cuando:

Sus datos personales sean objeto de un tratamiento automatizado, el cual le produzca efectos jurídicos no deseados o afecte de manera significativa sus intereses, derechos o libertades, y estén destinados a evaluar, sin intervención humana, determinados aspectos personales del mismo o analizar o predecir, en particular, su rendimiento profesional, situación económica, estado de salud, preferencias sexuales, fiabilidad o comportamiento.¹¹⁸

Esta disposición también es fundamental para hacer frente al ámbito de la discriminación algorítmica; sin embargo, debemos señalar sus limitaciones porque dentro de los sujetos obligados en la LGPDPSO no se incluye al sector privado (artículo 1.º, párrafos quinto, sexto y séptimo). Para este caso debe acudir a la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP) que, con todo, no prevé una disposición como la anterior.

Otras leyes a la que debe hacerse mención son la Ley Federal del Derecho de Autor (LFDA) y la Ley Federal de Protección a la Propiedad Industrial (LFPPI). En la primera de ellas se reconoce que los programas de cómputo (que incluyen a los algoritmos)¹¹⁹ son obras protegidas por el derecho de autor (artículo 13, fracción XI).

De manera tal que este tipo de obras están protegidas y, entre otros aspectos, no se autorizan intromisiones no consentidas en la información de tales sistemas informáticos. De hecho, el artículo 102 expresa que la protección se extiende al “código fuente” y el artículo 106, fracción IV, anticipa que, como parte del derecho patrimonial del titular de los derechos del programa, éste podrá autorizar o prohibir “los procesos para revertir la ingeniería de un programa de computación”, es decir, aquellos procesos que pueden llevar a conocer cómo es que el programa funciona.

¹¹⁸ Art. 47, fracción II, Ley General de Protección de Datos en Posesión de Sujetos Obligados, disponible en <https://www.diputados.gob.mx/LeyesBiblio/pdf/LGPDPSO.pdf> (acceso el 20 de junio de 2024).

¹¹⁹ De hecho, el artículo 101 define qué es un programa de computación, pero en su descripción incorpora múltiples características de los algoritmos, se indica: “Se entiende por programa de computación la expresión original en cualquier forma, lenguaje o código, de un conjunto de instrucciones que, con una secuencia, estructura y organización determinada, tiene como propósito que una computadora o dispositivo realice una tarea o función específica”, Artículo 101, Ley Federal del Derecho de Autor, disponible en <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFDA.pdf> (acceso el 20 de junio de 2024).

A su vez, la LFPPI expresa en su artículo 163, fracción I, que el secreto industrial es:

[...] toda información de aplicación industrial o comercial que guarde la persona que ejerce su control legal con carácter confidencial, que signifique la obtención o el mantenimiento de una ventaja competitiva o económica frente a terceros en la realización de actividades económicas y respecto de la cual haya adoptado los medios o sistemas suficientes para preservar su confidencialidad y el acceso restringido a la misma.¹²⁰

En esta hipótesis, a pesar de que no se hace referencia al programa informático o a sus algoritmos, la descripción del secreto industrial encaja con las características algorítmicas en tanto que es una información destinada al uso comercial, que se presume confidencial, que brinda una ventaja competitiva económica y que, por regla general, su propietario dispone los medios o sistemas suficientes para mantenerla en confidencialidad.

Ahora bien, en el ámbito jurisdiccional, nuestros tribunales han recurrido en ciertas ocasiones a las categorías de discriminación directa e indirecta. En la Tesis 1a. CCCLXXIV/2014, la Primera Sala de la Suprema Corte de Justicia de la Nación (SCJN) expresó que del artículo 1.º de la CPEUM se deriva la prohibición de discriminar, la cual puede ocurrir de manera indirecta “cuando una disposición, criterio o práctica aparentemente neutral ubica a un grupo social específico en clara desventaja frente al resto”.¹²¹ Además, expresó que la discriminación indirecta, a efectos de ser considerada en un juicio, debe permitir una comparación entre los diversos grupos involucrados y que debe “acreditarse empíricamente la afectación o desventaja producida en relación con los demás”,¹²² así como demostrar que ese trato no está amparado por algún “fin necesario”.¹²³ Es decir, si bien se reconoce la posibilidad de aplicar un criterio de discriminación indirecta, a juicio de nuestro máximo tribunal, es necesario

¹²⁰ Artículo 163, fracción I, Ley Federal de Protección a la Propiedad Industrial, disponible en <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPPI.pdf> (acceso el 20 de junio de 2024).

¹²¹ DISCRIMINACIÓN INDIRECTA O POR RESULTADOS. ELEMENTOS QUE LA CONFIGURAN. *Gaceta del Semanario Judicial de la Federación*, Décima Época, Libro 48, tomo I, noviembre de 2017, p. 225. Reg. Digital: 2015597, <https://sjf2.scjn.gob.mx/detalle/tesis/2015597>.

¹²² *Ibid.*

¹²³ *Ibid.*

acreditar de manera empírica el trato diferencial, así como la ausencia de toda forma de justificación legítima.¹²⁴

Estos criterios se contrastan con los emitidos por la Corte Interamericana de Derechos Humanos (Corte IDH) que también se ha pronunciado sobre la discriminación indirecta. En el caso “Pueblos Indígenas Maya Kaqchikel de Sumpango y otros vs. Guatemala”, la Corte señaló que el trato discriminatorio no sólo puede presentarse de manera deliberada, sino cuando se emplean categorías aparentemente neutras que generan efectos diferenciales en determinados sujetos.

En estos casos, para la Corte de San José, la discriminación se presenta per se, aun cuando no se pueda probar la intención de discriminar.¹²⁵ Igualmente, en el caso “San Miguel Sosa y otras vs. Venezuela”, la Corte expresó que, en determinadas situaciones de discriminación estructural, no “puede exigirse al empleador la prueba diabólica del hecho negativo de la discriminación”,¹²⁶ puesto que, “en este tipo de casos es prácticamente imposible para el recurrente demostrar fehacientemente un nexo causal, con pruebas directas, entre un trato discriminatorio y la decisión formal de terminar los contratos”.¹²⁷

Esta última situación resulta de gran interés para el caso de la discriminación algorítmica y, conforme a los parámetros de regularidad constitucional habría de ser una directiva por incorporar dentro de nuestra legislación.

124 Aspectos reiterados en diversas tesis, entre otras: DISCRIMINACIÓN INDIRECTA. SU DETERMINACIÓN REQUIERE DE UN ESTUDIO SOBRE LA EXISTENCIA DE FACTORES ESTRUCTURALES O CONTEXTUALES. Gaceta del Semanario Judicial de la Federación, Décima Época, Libro 34, tomo I, septiembre de 2016, p. 254. Reg. Digital: 2012596, <https://sjf2.scjn.gob.mx/detalle/tesis/2012596>; PENSIÓN COMPENSATORIA. LA PORCIÓN NORMATIVA “EN TANTO VIVA HONESTAMENTE Y NO CONTRAIGA NUPCIAS”, PREVISTA EN EL ARTÍCULO 162 DEL CÓDIGO CIVIL PARA EL ESTADO DE VERACRUZ, GENERA DISCRIMINACIÓN INDIRECTA POR RAZÓN DE GÉNERO, POR LO QUE ES INCOMPATIBLE CON EL ARTÍCULO 17 DE LA CONVENCIÓN AMERICANA SOBRE DERECHOS HUMANOS. Gaceta del Semanario Judicial de la Federación, Décima Época, Libro 54, tomo III, mayo de 2018, p. 2698. Reg. Digital: 2016940, <https://sjf2.scjn.gob.mx/detalle/tesis/2016940>; DISCRIMINACIÓN INDIRECTA O NO EXPLÍCITA. SU DETERMINACIÓN REQUIERE EL ANÁLISIS DE FACTORES CONTEXTUALES Y ESTRUCTURALES. Gaceta del Semanario Judicial de la Federación, Décima Época, Libro 58, tomo I, septiembre de 2018, p. 841. Reg. Digital: 2017989, <https://sjf2.scjn.gob.mx/detalle/tesis/2017989>.

125 Corte IDH, Caso Pueblos Indígenas Maya Kaqchikel de Sumpango y otros Vs. Guatemala, “Fondo, Reparaciones y Costas. Sentencia de 6 de octubre de 2021”, Serie C No. 440, párr. 136, http://www.corteidh.or.cr/docs/casos/articulos/seriec_440_esp.pdf

126 Corte IDH. Caso San Miguel Sosa y otras Vs. Venezuela, “Fondo, Reparaciones y Costas. Sentencia de 8 de febrero de 2018”, Serie C No. 348, párr. 192, http://www.corteidh.or.cr/docs/casos/articulos/seriec_348_esp.pdf

127 *Ibid.*

Como colofón de este apartado, podemos observar que, salvo algunas excepciones, el panorama que se nos presenta en el derecho interno para hacer frente al fenómeno de la discriminación algorítmica resulta limitado. En efecto, formas de discriminación directa podrían ser atajadas por las actuales disposiciones, puesto que bastaría acreditar que nos encontramos ante una categoría sospechosa y que el trato no ha sido racional, objetivo y proporcional. Sin embargo, la discriminación que en mayor medida ocurre en el caso de los algoritmos es la que se produce de manera sutil y por efecto o consecuencia diferencial, es decir, las formas de discriminación indirecta. Sin embargo, nuestra legislación se posiciona preferentemente por hipótesis de discriminación directa y, en aquellos pocos casos donde se plantea la discriminación indirecta se requiere acreditar el nexo causal, cuestión que también resulta compleja porque, como se recordará, un algoritmo puede trabajar con proxys o ser discriminatorio mediante variables sumamente veladas, además, tampoco se puede requerir al patrón, programador o empresario que proporcione el código fuente porque éste se encuentra protegido también en nuestra legislación, ya sea por derechos de autor o por el secreto industrial. Además, aunque en materia de protección de datos existen algunos avances, su aplicabilidad se restringe al ámbito público.

V. REFLEXIONES DESDE LA ÓPTICA LEGISLATIVA. RETOS Y OPORTUNIDADES

Visto el escenario actual en nuestra legislación, veamos qué herramientas o buenas prácticas podrían ser incorporadas en una perspectiva futura (lege ferenda) para hacer frente al fenómeno de la discriminación laboral algorítmica.

Una primera alternativa pasa por la modificación de las disposiciones dirigidas a prevenir la discriminación para que, tal y como lo hace la LFPED, se reconozca no sólo las formas de discriminación que tienen origen o se motivan por alguna categoría sospechosa (los inputs), sino también aquellas formas de discriminación en donde lo relevante no es el motivo, sino el efecto o la consecuencia (el output), es decir, la discriminación indirecta. Además, en estos casos es importante, tal y como lo señala la Corte IDH, que la carga de la prueba no se condicione

al demandante porque se trata de condiciones estructurales presentes al interno del propio sistema y que resultan difíciles de demostrar para las personas que se ven afectadas, por tal motivo debería bastar el efecto o resultado sin que sea necesario acreditar la relación causal (que, por ejemplo, hemos ya visto es casi imposible de demostrar en algunos tipos de algoritmos).

Asimismo, puede acudir a reflexiones de derecho comparado donde se observan regulaciones en la materia. Al respecto, es de gran interés lo que ocurre en el ámbito comunitario europeo, particularmente derivado de la aprobación en marzo de 2024 del Reglamento de IA, pero, incluso, ya desde 2016 se disponía de algunas normas al respecto. El ejemplo paradigmático es el Reglamento General de Protección de Datos de la Unión Europea (UE) que en su artículo 22 instituye el derecho de toda persona a no ser objeto de decisiones basadas únicamente en el tratamiento automatizado cuando de ello puedan derivarse efectos jurídicos sobre el titular de esos datos (salvo algunas excepciones). El reglamento expresa que esto incluye la elaboración de perfiles (artículo 4.o, numeral 4), es decir:

[...] toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales de una persona física, en particular para analizar o predecir aspectos relativos al rendimiento profesional, situación económica, salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos de dicha persona física.¹²⁸

Como se observa, dicha disposición se asemeja a la presente en la LGPDPSO, sin embargo, sus efectos resultan de mucho mayor calado puesto que el Reglamento europeo, en términos de su artículo 2.o, se aplica sin distinción al ámbito público y privado dentro de la UE.

El derecho a no ser objeto de decisiones que se basen únicamente en tratamientos automatizados viene a justificarse en tanto que, como hemos visto, una decisión de este tipo puede condicionar aspectos como el empleo, el acceso a un crédito hipotecario o incluso la libertad, es decir, puede proyectarse en derechos fundamentales y, por tanto, este tipo

¹²⁸ Art. 4.o, numeral 4, Reglamento General de Protección de Datos, 27 de abril de 2016, <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32016R0679> (acceso el 20 de junio de 2024).0

de decisiones no deben hacerse depender de un sistema automatizado que sólo instrumenta al individuo y en donde ni siquiera existe certeza respecto de cuáles han sido los parámetros de su funcionamiento.¹²⁹

Por esta razón, además de prohibirse la toma de decisiones basadas exclusivamente en el tratamiento automatizado, también se incorporan garantías adicionales, como la necesaria intervención humana (artículo 14), la posibilidad de recurrir la decisión automatizada (artículo 22, párrafo tercero) e, incluso, el derecho a una “explicabilidad” sobre esa decisión (artículos 13, numeral 3, inciso b, IV, y 86).

En este último caso, se ha reconocido que el afectado tiene el derecho a recibir una explicación sobre la lógica subyacente en la tecnología que se le ha aplicado y, esta explicación debe ser clara y significativa, de manera tal que no pueden admitirse respuestas del tipo “entrenamos el sistema usando una enorme cantidad de datos, y esto es lo que decidió”.¹³⁰ Todas estas medidas resultan fundamentales para hacer frente a la opacidad y arbitrariedad que pueden derivarse del efecto “caja negra”.¹³¹

La normativa anterior se complementa con el Reglamento de IA que, entre muchas otras cuestiones, clasifica esta tecnología con base en sus riesgos potenciales. Atendiendo a cada nivel de riesgo dispone correlativas medidas de precaución. Curiosamente, el empleo de IA en el ámbito laboral es considerado un riesgo “alto”, de tal manera que no está prohibido, pero sí sujeto a rigurosos controles. Más aún, expresamente se reconoce el caso de la aplicación de sistemas de IA utilizados “para publicar anuncios de empleo específicos, analizar y filtrar las solicitudes de empleo y evaluar a los candidatos”.¹³²

Dentro de las obligaciones que se prevén para el caso de tecnología de IA con riesgo alto se dispone que, en cada producto y en cada proceso de creación, programación y aplicación deberá garantizarse que los datos

¹²⁹ Falletti, “Algorithmi: la discriminazione non...”, 9; Capuzzo, “Do Algorithms dream about Electric Sheep?...”, 99-100.

¹³⁰ Tegmark, *Vida 3.0. Ser humano en...*, 124.

¹³¹ Ramírez Bustamante y Páez, “Análisis jurídico de la discriminación algorítmica...”, 19.

¹³² El artículo 6.o, numeral 2, del Reglamento señala en qué consiste el nivel de riesgo alto. Además, prevé diversas hipótesis en las cuales este riesgo se actualiza, para ello reenvía al Anexo III del Reglamento. En dicho anexo el ámbito laboral se encuentra previsto en el numeral 4. De igual modo, se destaca que el artículo 6.o del Reglamento advierte que los sistemas de IA “siempre se considerarán de alto riesgo cuando el sistema de IA lleve a cabo la elaboración de perfiles de personas físicas”. Artículos 4.o, numeral 4, y 6.o, numeral 2, Reglamento de Inteligencia Artificial, 13 de marzo de 2024, https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_ES.pdf (acceso el 20 de junio de 2024).

sean de calidad, representativos, libres de errores o sesgos y completos (artículo 8.o), que deberá seguirse un adecuado proceso de trazabilidad y documentación en todas las etapas (artículo 9.o), que deberá cumplirse con las obligaciones en materia de transparencia (artículo 10), que se garantizará la exactitud y seguridad de la tecnología (artículo 12) y que deberá contarse con la supervisión humana (artículo 11).¹³³

Además, y de manera coherente con el principio de precaución, el Reglamento dispone que los responsables de sistemas de IA de alto riesgo habrán de realizar, previo al despliegue del sistema, una “evaluación de impacto relativa a los derechos fundamentales” (artículo 27), donde se incluirá, entre otros elementos: las personas o grupos que pueden verse afectados, los riesgos de perjuicio a derechos fundamentales y las medidas diversas que han de implementarse para minimizarlos. A la par, la evaluación contendrá las medidas de gobernanza interna y los mecanismos de reclamación habilitados para que los potenciales afectados ejerzan su derecho a la defensa. Dicha evaluación resulta crucial porque en ella pueden identificarse, y corregirse, las deficiencias, sesgos o errores del programa y atenderlas antes de que pueda aplicarse a casos concretos con potenciales efectos en los derechos fundamentales.¹³⁴

Por cierto, este procedimiento es independiente, aunque puede complementarse, con la evaluación de impacto que también se prevé en el Reglamento General de Protección de Datos (artículo 35, párrafo primero), al indicar que:

Quando sea probable que un tipo de tratamiento, en particular si utiliza nuevas tecnologías, por su naturaleza, alcance, contexto o fines, entrañe un alto riesgo para los derechos y libertades de las personas físicas, el responsable del tratamiento realizará, antes del tratamiento, una evaluación del impacto de las operaciones de tratamiento en la protección de datos personales [...].¹³⁵

El mismo artículo dispone en su numeral 3 que la evaluación de impacto se realizará, en particular, tratándose de los sistemas de tratamiento

¹³³ Véase en detalle: Alameda Castillo, “Reclutamiento tecnológico. Sobre algoritmos...”, 41-42; Sáez Lara, “El algoritmo como protagonista de la relación laboral...”, 51-57.

¹³⁴ Mantelero, “A blueprint for a human right, social and ethical impact assessment”, 772.

¹³⁵ Artículo 35, párrafo primero, Reglamento General de Protección de Datos.

automatizado, la elaboración de perfiles y el tratamiento a gran escala de datos. Al igual que como ocurre en el Reglamento de IA, también aquí deben identificarse los potenciales riesgos para la protección de datos e incorporarse los controles conducentes para evitarlos (artículo 35, párrafo séptimo).

A su vez, tanto en materia de protección de datos como específicamente en temas de IA, los reglamentos disponen un sistema para la determinación de responsabilidades. En estos casos, los reglamentos optan por dar primacía al principio de precaución que se sitúa en el ámbito *ex ante* de este tipo de sistemas. Esta particularidad permite que el sistema de responsabilidad se asiente en la incorporación de las medidas preventivas dirigidas a evitar la afectación de los derechos fundamentales, entre ellas la evaluación de impacto. De esta manera, incluso en aquellos casos en los que pudiera pensarse en un sistema de IA con autonomía decisional (de Machine Learning) o cuando la discriminación fuese involuntaria, cabría derivar una clara responsabilidad. Esta responsabilidad podría ser el resultado de la omisión de haber efectuado el correspondiente estudio de impacto o de haberlo realizado de manera indebida, no anticipando riesgos que eventualmente podrían derivarse (por negligencia, impericia o falta de cuidado).¹³⁶

Como puede advertirse, el ámbito comunitario europeo dispone ya de algunas reglas dirigidas a prevenir la afectación a los derechos fundamentales, incluyendo la no discriminación que pueden derivarse de las nuevas tecnologías. Llegados a este punto y dadas las limitaciones de esta investigación es momento de detenernos aquí, dejando reflexiones de mayor profundidad para próximos trabajos.

VI. CONCLUSIONES

Asistimos a una nueva revolución tecnológica representada por la IA y sus múltiples aplicaciones. Esta tecnología puede ser empleada para el bienestar de la humanidad, pero también de manera contraria. Por esta razón es importante disponer de marcos jurídicos claros bajo los cuales

¹³⁶ Sáez Lara, “El algoritmo como protagonista de la relación laboral...”, 51; Alameda Castillo, “Reclutamiento tecnológico. Sobre algoritmos...”, 44-46; Rivas Vallejo, *Discriminación algorítmica...*, 53.

podamos aprovechar sus beneficios y anticiparnos a sus peligros, el ámbito de los derechos humanos no está exento de ello y, tanto menos aquellos derechos que se vinculan al ámbito laboral.

Hoy en día múltiples empresas emplean algoritmos informáticos a los que han incorporado tecnología de IA. Estas herramientas permiten la optimización de los recursos, incluyendo aquellos humanos. De esta manera, el empleador puede aplicarlos para medir la productividad y eficiencia y, a partir de ello, derivar decisiones empresariales como podrían ser la continuidad o el despido del trabajador. Esta tecnología es susceptible de ser empleada en toda la etapa laboral, sin embargo, nos hemos concentrado en la fase de contratación donde se han identificado diversas hipótesis que podrían dar lugar a formas de discriminación.

La tecnología algorítmica no es discriminatoria por sí misma, su grado de objetividad y certeza dependerá de diversos factores, pero sobre todo de la información que se emplea. Si la información es de buena calidad, los resultados también lo serán, de lo contrario podrán derivarse decisiones alejadas de la realidad o bien, coherentes con la realidad, pero con potenciales efectos discriminatorios por los que se termine dando continuidad a históricas desigualdades existentes en nuestras sociedades.

Para evitar lo anterior, los datos deben ser: objetivos, limpios, cuantitativamente significativos, disponer de la adecuada representatividad de grupos en situación de vulnerabilidad, actuales, ausentes de toda categoría sospechosa o que pueda vincularse a ella mediante el empleo de variables de proximidad y, además, debe cuidarse de las subjetividades, sesgos o experiencias personales de la persona responsable de dar vida al algoritmo (tanto el programador como el empleador o propietario). En temas de derechos humanos, también habría de configurar y entrenar al algoritmo con aspectos como la perspectiva de género, el principio pro persona, la necesidad de incorporar acciones afirmativas y otros tópicos que hoy en día asumimos como principios inquebrantables.

En el ámbito normativo nuestra legislación dispone de ciertas normas de principio que podrían ser empleadas al ámbito de la discriminación algorítmica, sin embargo, son pocas y en muchos casos inadecuadas para hacer frente a formas de discriminación indirecta o por resultado puesto que, en la mayor parte de los casos se requiere acreditar que el tratamiento diferenciado ha sido el resultado o motivo de alguna de las categorías sospechosas. La discriminación algorítmica suele vincularse

con categorías que en apariencia son neutras pero cuya aplicación genera efectos diferenciales en determinadas categorías de sujetos, normalmente en situación de vulnerabilidad, además, el algoritmo suele estar protegido por derechos de autor y secreto empresarial, de manera tal que no puede saberse con precisión cuáles han sido las variables o criterios empleados por el algoritmo para tomar su decisión, aunado al hecho de que cuando se trata de algoritmos de Machine Learning puede resultar imposible conocer su funcionamiento, aun para el propio programador o empresario. Por esta razón, habría de acudir a una legislación que permitiera combatir la discriminación por sus efectos o resultados y sin que resultara determinante acreditar la relación causal, proceder a que es respaldado incluso por la Corte IDH (y del cual nuestros tribunales se alejan).

Asimismo, en perspectiva de lege ferenda, resultaría adecuado actualizar nuestra normativa para que, o bien, pudiera incluirse directamente la categoría de la discriminación algorítmica o bien, actualizarla para poder responder con eficacia a las diversas hipótesis de discriminación por efectos o resultados. A la par, sería conveniente replicar disposiciones como la presente en el artículo 47, fracción II, de la LGPDPSO (que guarda coincidencia con el artículo 22 del Reglamento europeo en materia de protección de datos) al ámbito privado donde las hipótesis de discriminación algorítmica suelen presentarse con mayor intensidad.

Igualmente, disposiciones como la anterior podrían fortalecerse con el reconocimiento del derecho a impugnar decisiones automatizadas, así como a requerir la explicabilidad en el funcionamiento de los algoritmos, aspectos basilares que permitirían evitar el efecto de “caja negra” y, en consecuencia, la opacidad y arbitrariedad de este tipo de decisiones. Además, el derecho a la explicación podría compatibilizarse con el secreto empresarial y los derechos de autor puesto que ello no supone el acceso al código fuente, sino una explicación que, asentada en elementos objetivos, permita entender de manera razonada las variables y el funcionamiento que han sido empleadas para derivar de ciertos datos de entrada, determinados datos de salida.

A la par, a fin de afrontar los efectos futuros de una previa perfilación laboral, resultaría conveniente fortalecer los derechos de acceso, rectificación, oposición y cancelación en el ámbito de la tutela de los datos para, además de informar en todo momento que será empleado un tratamiento automatizado, se dispongan las garantías para que los

candidatos descartados (y aún aquellos seleccionados) puedan saber el destino y suerte de su información. Igualmente, el titular de estos datos habría de saber en todo momento quién mantiene esa información y, si lo desea, requerir que su información sea retirada de una eventual perfilación, inclusive, por la relevancia del tema, disposiciones de este tipo podrían ser incorporadas en el ámbito de la legislación laboral.

Otra buena práctica que podría implementarse en nuestras leyes es la de disponer evaluaciones de impacto en el empleo de este tipo de tecnologías. Al reconocerse que un uso inadecuado de la tecnología de IA es susceptible de afectar derechos humanos, resulta coherente el planteamiento de elaborar estudios de impacto cuando algoritmos con este tipo de tecnología son empleados en la toma de decisiones con incidencia en los derechos humanos. De esta manera, la tecnología podría aplicarse siempre que hubiere superado el referido estudio en el que, como parte de sus requisitos, habría de disponerse el análisis de riesgo y las respectivas medidas dirigidas a prevenirlos. Este estudio podría complementarse con una suerte de certificación, inspección o vigilancia a cargo de la autoridad laboral o, inclusive, de ésta en coordinación de las autoridades competentes en materia de protección de datos y de desarrollo tecnológico.

Finalmente, no debemos perder de vista que los algoritmos y la tecnología de IA hacen parte de nuestra vida cotidiana, resultaría iluso detener su avance o plantear toda forma de prohibición, lo que sí puede y debe hacerse es, como ya señalamos, identificar los posibles riesgos de esta tecnología e incorporar las medidas conducentes, incluyendo las de orden legislativo, que permitan compatibilizar de manera eficaz el ejercicio de los múltiples intereses en juego, sin perder de vista que, tratándose de derechos humanos, existe un deber de prevalencia al que todos nuestros objetivos habrían de estar dirigidos.

VII. REFERENCIAS

1. **Bibliohemerográficas**

Abadías Selma, Alfredo. *Child Grooming: el embaucamiento de menores en la era del metaverso y la inteligencia artificial*. Madrid: Editorial Tecnos, 2024.

- . El derecho penal frente a la discriminación laboral algorítmica. Navarra: Editorial Aranzadi, 2023.
- . Justicia juvenil e inteligencia artificial en la era de la cultura “Touch”. Valencia: Editorial Tirant Lo Blanch, 2022.
- Alameda Castillo, María Teresa. “Reclutamiento tecnológico. Sobre algoritmos y acceso al empleo”. *Temas laborales. Revista andaluza de trabajo y bienestar social (Sevilla)*, n.o 159 (2021), 11-52. <https://dialnet.unirioja.es/servlet/articulo?codigo=8250088>
- Alexy, Robert. *Teoría de los derechos fundamentales*. 2ª ed. Traducido por Carlos Bernal Pulido. Madrid: Centro de Estudios Constitucionales, 2012.
- . Tres escritos sobre derechos fundamentales y la teoría de los principios. Traducido por Carlos Bernal Pulido. Bogotá: Universidad Externado de Colombia, 2003.
- Ali, Muhammad, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove y Aaron Rieke. “Discrimination through optimization: How Facebook’s ad delivery can lead to skewed outcomes”. *ArXiv (Nueva York)*, (12 septiembre, 2019), 1-17. <https://doi.org/10.1145/3359301>
- Asquerino Lamparero, María José. “Algoritmos y discriminación”, en José Morales Ortega (coord.). *Realidad social y discriminación: Estudios sobre diversidad e inclusión laboral*. Murcia: editorial Laborum, (2022), 335-382. <https://idus.us.es/handle/11441/133121>
- . “Algoritmos, procesos de selección y reputación digital: una mirada antidiscriminatoria”. *Documentación Laboral (Madrid)*, n.o 126 (2022), 121-136. <https://dialnet.unirioja.es/servlet/articulo?codigo=8653047>.
- Barbera, Marzia. “Discriminazioni algoritmiche e forme di discriminazione”. *Labour & Law Issues (Bologna)* 7, n.o 1 (2021), 1-17. doi: <https://doi.org/10.6092/issn.2421-2695/13127>
- Barocas, Solon y Andrew Selbst. “Big data’s disparate impact”. *California Law Review (Berkeley)* 104, n.o 3 (2016), 671-732. <https://www.jstor.org/stable/i24758717>
- Basile, Fabio. “Intelligenza artificiale e diritto penale quattro possibili percorsi di indagine”. *Diritto penale e uomo (Milán)*. (septiembre 2019), 1-33. <https://archiviodpc.dirittopenaleuomo.org/upload/3089-basile2019.pdf>

- Bogen, Miranda y Aaron Rieke. Help wanted. an examination of hiring algorithms, equity and bias. Washington DC: editorial Upturn, 2018.
- Bostrom, Nick. Superinteligencia. Caminos, peligros, estrategias. 3ª ed. Traducido por Marcos Alonso. Zaragoza: Teell editorial, 2016.
- Capuzzo, Giacomo. “Do Algorithms dream about Electric Sheep? Percorsi di studio in tema di discriminazione e processi decisori algoritmici tra le due sponde dell’Atlantico”. Media Laws. Rivista di Diritto dei Media (Milán), n.o 3, (octubre 2020), 89-106. https://www.medialaws.eu/rivista_category/3-2020/.
- Caterini, Mario. “El sistema penal en la encrucijada ante el reto de la inteligencia artificial”. Revista de Internet, Derecho y Política (Barcelona) 35, (marzo 2022), 1-19. <https://dialnet.unirioja.es/descarga/articulo/8399208.pdf>.
- Datta, Amit, Michael Carl Tschantz y Anupam Datta. “Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination”. ArXiv (Nueva York), (18 marzo 2015), 1-26. <https://arxiv.org/abs/1408.6491>.
- De Vicente Pachés, Fernando. “Acoso y discriminación en el trabajo a través de dispositivos tecnológicos o digitales: su incidencia en la fase de selección y acceso al empleo”, en Sara Ruana Albertos (coordinadora). Nuevas formas de acceso al empleo. Barcelona: editorial Atelier, (2019), 181-222. doi: <https://doi.org/10.5944/rdpc.20.2018.26446>
- Domingos, Pedro. The Master Algorithm. Nueva York: editorial Basic Books, 2018.
- Eguiluz Castañeira, Josu. “Desafíos y retos que plantean las decisiones automatizadas y los perfilados para los derechos fundamentales”. Estudios de Deusto (Bilbao) 68, n.o 2 (2020), 325-367. doi: [https://doi.org/10.18543/ed-68\(2\)-2020pp325-367](https://doi.org/10.18543/ed-68(2)-2020pp325-367)
- Esposito, Andreana. “Note sparse sull’intelligenza artificiale”, en Giuliano Balvi, Federica De Simone, Andreana Esposito y Stefano Manacorda (coords.). Diritto penale e intelligenza artificiale, “Nuovi scenari”. Materiali e Studi di Diritto Pubblico. Turín: Editorial Giappichelli, 2022, 37-49. <https://www.giappichelli.it/media/catalog/product/openaccess/9788892177710.pdf>
- Falletti, Elena. “Algoritmi: la discriminazione non è uguale per tutti”. Lavoro, diritti, Europa. Rivista nuova di Diritto del Lavoro (Tu-

- rín), n.o 2 (2023), 2-14. https://www.lavorodirittieuropa.it/images/FALLETTI_2.pdf
- Fernández García, Antonio. “La lucha contra la discriminación en el acceso al empleo mediante el currículum vitae anónimo”. *IUSlabor. Revista d’anàlisi de Dret del Treball* (Barcelona), n.o 1 (2019), 233-266. <https://doi.org/10.31009/IUSLabor.2019.i01.05>.
- Ferrajoli, Luigi. *Derechos y garantías. La ley del más débil*. 7ª ed. Traducido por Perfecto Andrés Ibáñez. Madrid: editorial Trotta, 2010.
- Ferrante, Enzo. “Inteligencia artificial y sesgos algorítmicos. ¿Por qué deberían importarnos?”. *Nueva Sociedad* (Santa Fe), n.o 294 (julio–agosto 2020), 27-37. https://static.nuso.org/media/articles/downloads/1.TC_Ferrante_294.pdf
- Fita Ortega, Fernando. “El derecho a la mentira como tutela preventiva de los derechos fundamentales de los trabajadores”. *Revista de Trabajo y Seguridad Social*-I. Centro de Estudios Financieros (Madrid), n.o 441, (diciembre 2019), 85-116. doi: <https://doi.org/10.51302/rtss.2019.1166>
- Flores, Juan y Félix Calderón. “Procesos de decisión de Markov y aprendizaje por refuerzo”, en Grigori Sidorov, *Métodos modernos de Inteligencia Artificial*. Ciudad de México: Sociedad Mexicana de Inteligencia Artificial, 2011, 357-405.
- Gaudio, Giovanni. “Le discriminazioni algoritmiche”. *Lavoro, diritti, Europa. Rivista nuova di Diritto del Lavoro* (Turín), n.o 1 (2024), 1-26. <https://www.lavorodirittieuropa.it/images/GAUDIO.pdf>
- Habermas, Jürgen. *Facticidad y validez. Sobre el derecho y el Estado democrático de derecho en términos de teoría del discurso*. 6ª ed. Traducido por Manuel Jiménez Redondo. Madrid: editorial Trotta, 2010.
- Huberman, Leo. *Los bienes terrenales del hombre*. Editado por Gabriel Silva Rincón. Bogotá: editorial Panamericana, 2007.
- Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan y Cass R. Sunstein. “Discrimination in the Age of Algorithms”. *Journal of Legal Analysis* (Oxford), n.o 10 (2018), 113-174. <https://doi.org/10.1093/jla/laz001>
- Kosinski, Michal, David Stillwell y Thore Graepel. “Private traits and attributes are predictable from digital records of human behavior”. *Proceedings of the National Academy of Sciences* (Washington) 110, n.o 15 (abril 2013), 5802-5805. <https://www.pnas.org/doi/pdf/10.1073/pnas.1218772110>

- Maldonado Smith, Mario Eduardo. “Reflexiones iuspenalistas sobre los delitos contra la libertad sindical en la era digital”, en Alfredo Abadías Selma y Guillermo García González, *Protección de los trabajadores e inteligencia artificial: La tutela de los derechos sociales en la cuarta revolución industrial*. Barcelona: editorial Atelier, 2022, 261-286.
- Mantelero, Alessandro. “AI and Big Data: A blueprint for a human right, social and ethical impact assessment”. *Computer Law and Security Review (Amsterdam)* 34, n.o 4, (agosto 2018), 754-772. <https://www.sciencedirect.com/science/article/pii/S0267364918302012>
- Marcuzzolo Awad, Ximena y Nathalie Walker Silva. “Empleo de sistemas algorítmicos de evaluación de riesgos en materia penal. Estándares mínimos para un uso acorde a las exigencias del debido proceso”, en Martín Rocha Espíndola, Daniel Sansó, Rubert Pascual y Nuria Rodríguez Dos Santos (coord.). *Inteligencia artificial y derecho. Reflexiones jurídicas para el debate sobre su desarrollo y aplicación*. Madrid: Dykinson, 2023.
- Mercader Uguina, Jesús. “Algoritmos y derecho del trabajo”. *Actualidad Jurídica Uría Menéndez (Madrid)*, n.o 52 (junio–septiembre 2019), 63-70. <https://www.uria.com/es/revista/58>
- . “Discriminación algorítmica y derecho granular: nuevos retos para la igualdad en la era del big data”. *Labos. Revista de Derecho del Trabajo y Protección Social (Madrid)* 2, n.o 2 (julio 2021), 4-11. doi: <https://doi.org/10.20318/labos.2021.6213>
- Merino Rus, Rafael. *La igualdad en la era de los algoritmos y la inteligencia artificial*. Bizkaia: Gernika Gogoratzu, 2020.
- Miró Llinares, Fernando. “Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots”. *Revista de Derecho Penal y Criminología (Madrid)* 20, (julio 2018), 87-130.
- Moccia, Sergio. “Sistema penale e principi costituzionali: un binomio inscindibile per lo stato sociale di diritto. Relazione di sintesi”. *Rivista Italiana di Diritto e Procedura Penale (Milán)* 61, n.o 3 (2018), 1720-1753.
- Molina Navarrete, Cristóbal. “Duelo al sol. ¿Un algoritmo controla mi trabajo? Sí; a tu empresa también”. *Revista de Trabajo y Seguridad Social. Centro de Estudios Financieros (Madrid)*, n.o 457 (abril 2021), 5-21. doi: <https://doi.org/10.51302/rtss.2021.2370>

- Molina, Óscar Andrés. “Regulación de la Inteligencia Artificial aplicada a la medicina: actualidad y propuestas”, en Martín Rocha Espíndola, Daniel Sansó, Rubert Pascual y Nuria Rodríguez Dos Santos (coord.). *Inteligencia artificial y derecho. Reflexiones jurídicas para el debate sobre su desarrollo y aplicación*. Madrid: Dykinson, 2023.
- Morondo Taramundi, Dolores. “Le sfide della discriminazione algoritmica”. *Genius. Rivista di studi giuridici sull’orientamento sessuale e identità di genere* (Turín) 21 (diciembre 2022), 22-32.
- Mueller, John y Luca Massaron. *Machine Learning for dummies*. Nueva York: editorial John Wiley & Sons, 2016.
- Nino, Carlos Santiago. *Ética y derechos humanos. Un ensayo de fundamentación*. Barcelona: editorial Astrea, 1989.
- Núñez Zorrilla, María del Carmen. *Inteligencia artificial y responsabilidad civil. Régimen jurídico de los daños causados por robots autónomos con inteligencia artificial*. Madrid: editorial Reus, 2019.
- O’Neil, Cathy. *Weapons of math destruction. How Big Data increases inequality and threatens democracy*. New York: Editorial Crown, 2016.
- Olarte Encabo, Sofía. “La aplicación de la inteligencia artificial a los procesos de selección de personal y ofertas de empleo: impacto sobre el derecho a la no discriminación”. *Documentación Laboral* (Madrid), n.o 119 (2020), 79-98. <https://dialnet.unirioja.es/servlet/articulo?codigo=7544418>
- Pasquale, Frank. *The Black Box Society: The secret algorithms that control money and information*. Cambridge: Harvard University Press, 2015.
- Pinto, Mónica. *Temas de derechos humanos*. Buenos Aires: Editorial Universidad de Buenos Aires, 2009.
- Ramírez Autrán, Rodrigo. “Sesgos y discriminaciones sociales de los algoritmos en Inteligencia Artificial: una revisión documental”. *Entretextos* (Ciudad de México) 15, n.o 39 (2023), 1-17. <https://dialnet.unirioja.es/servlet/articulo?codigo=9380872>
- Ramírez Bustamante, Natalia y Andrés Páez. “Análisis jurídico de la discriminación algorítmica en los procesos de selección laboral”. *Social Science Research Network, SSRN* (Nueva York), (7 abril 2021), 1-29. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3765741

- Rifkin, Jeremy. *La sociedad del coste marginal cero. El internet de las cosas, el procomún colaborativo y el eclipse del capitalismo*. Traducido por Genís Sánchez Barberán. Paidós: Barcelona, 2014.
- Rivas Vallejo, Pilar. *Discriminación algorítmica: detección, prevención y tutela*. XXXI Jornadas Catalanes de Dret Social. Universidad de Barcelona: Barcelona, 2021. <http://www.iuslabor.org/wp-content/plugins/download-monitor/download.php?id=415>
- Rodotà, Stefano. *Tecnopolitica. La democrazia e le nuove tecnologie della comunicazione*. Roma-Bari: editorial Laterza, 1997.
- Sáez Lara, Carmen. “El algoritmo como protagonista de la relación laboral. Un análisis desde la perspectiva de la prohibición de discriminación”. *Temas laborales. Revista andaluza de trabajo y bienestar social* (Sevilla), n.o 155 (2020), 41-60. <https://dialnet.unirioja.es/servlet/articulo?codigo=7801689>
- Serna, Ignacio, Alejandro Peña, Aythami Morales y Julián Fierrez. “Inside Bias: Measuring bias in deep networks and application to face gender biometrics”. *ArXiv* (Nueva York), (22 julio 2020), 1-8. <https://doi.org/10.48550/arXiv.2004.06592>
- Simón Castellano, Pere. *Justicia cautelar e inteligencia artificial. La alternativa a los atávicos heurísticos judiciales*. Barcelona: editorial J.M. Bosch, 2021.
- Soria Olivas, Emilio, Pablo Rodríguez Belenguer, Enrique García Vidal, Fran Vaquer Estalrich, Juan Vicent Camisón y Jorge Vila Tomás. *Inteligencia Artificial. Casos prácticos con Aprendizaje Profundo*. Madrid: Ra-Ma editorial, 2022.
- Sweeney, Latanya. “Discrimination in Online Ad Delivery”. *ArXiv* (Nueva York), (28 enero 2013), 1-36. <https://arxiv.org/pdf/1301.6822>
- Tegmark, Max. *Vida 3.0. Ser humano en la era de la Inteligencia Artificial*. Traducido por Marcos Pérez Sánchez. Madrid: Taurus, 2018.
- Tommasi, Sara. “Algoritmo e nuove forme di discriminazione: uno sguardo al diritto europeo”. *Rivista de Direito Brasileira* (Florianopolis) 27, n.o 10, (septiembre - diciembre 2020), 112-129. <https://www.indexlaw.org/index.php/rdb/article/download/7293/5337>
- Urteaga, Eguzki. “Las políticas de discriminación positiva”. *Revista de Estudios Políticos* (Madrid) 146, (octubre – diciembre 2009), 181-213. <https://dialnet.unirioja.es/descarga/articulo/3091251.pdf>

- Vilares Ferro, Manuel, Santiago Fernández Lanza, Milagro Fernández Gavilanes y Francisco Ribadas Pena. “Ontologías y representación del conocimiento”, en Grigori Sidorov. Métodos modernos de Inteligencia Artificial. Ciudad de México: Sociedad Mexicana de Inteligencia Artificial, 2011, 84-146.
- Williams, Patricia. La dolorosa prisión del lenguaje de los derechos, en Wendy Brown y Patricia Williams. La crítica de los derechos. Bogotá: Siglo del Hombre Editores, 2003, 45-73.
- Xenidis, Raphaele y Linda Senden. “EU Non-Discrimination Law in the era of Artificial Intelligence: Mapping the challenges of algorithmic discrimination”, en Ulf Bernitz, Xavier Groussot, Jaan Paju y Sybe de Vries (coords.). General Principles of EU Law and the EU Digital Order. Londres: editorial Kluwer Law International, 2020, 151-182. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3529524

2. Otros

Carta Ética Europea sobre el uso de la Inteligencia Artificial en los sistemas judiciales y su entorno.