

## ARTÍCULOS



## ALGORITMO Y DISCRIMINACIÓN\*

### ALGORITHM AND DISCRIMINATION

### ALGORITHMHE ET DISCRIMINATION

---

Francisco Javier CALVO GALLEGOS\*\*

RESUMEN: El imparable uso de la Inteligencia Artificial supone ciertamente enormes ventajas y oportunidades, pero también muy importantes riesgos. El principal de ellos, y ya constatado, es el de ocultar o incluso intensificar discriminaciones. Este trabajo pretende analizar, a la espera de una posible regulación europea —anunciada una vez cerrado este trabajo—, sus causas, así como los problemas que plantea la posible aplicación normativa antidiscriminatoria y el importante papel al que está llamada la regulación europea de protección de datos. El estudio se cierra destacando la necesidad de una intervención multidimensional, que evite no sólo las posibles desigualdades y exclusiones sociales motivadas por su aplicación, sino que también garantice una IA segura, ética y robusta, con el apoyo de la responsabilidad social de las empresas.

*Palabras clave:* algoritmo, discriminación, inteligencia artificial.

ABSTRACT: The growing use of Artificial Intelligence around the world certainly produces enormous advantages and opportunities, but it also generates significant challenges. The main one, and already verified, is the

---

\* Recibido el 3 de mayo de 2020 y aceptado para su publicación el 15 de septiembre de 2021.

Este trabajo forma parte de los resultados del Proyecto de Investigación denominado Nuevas causas y perfiles de discriminación e instrumentos para la tutela antidiscriminatoria en el nuevo contexto tecnológico social (Proyecto I+D+i FEDER Andalucía; US: 1264479)

\*\* Catedrático del Departamento de Derecho del Trabajo y de la Seguridad Social de la Universidad de Sevilla. Doctor en Derecho por la Universidad de Bolonia (Italia) y Magíster en Derecho del Trabajo y de la Seguridad Social por la Universidad de Sevilla. Becario FPI, de la Fundación Oriol Urquijo y del Real Colegio de España en Bolonia.



risk of hiding or even intensifying discrimination. In anticipation of a possible European regulation —announced when this work was written—, this essay aims to analyse, on one hand, the factors that causes these biases, the anti-discrimination regulations response, and the important role that European data protection regulation must play in this field. The paper ends highlighting the need for a multidimensional intervention, which, not only avoids the possible inequalities and social exclusions motivated by the IA, but also guarantees a safe, ethical and robust AI, with the support of the social responsibility of enterprises.

*Key words:* algorithm, discrimination, artificial intelligence.

RÉSUMÉ: L'imparable utilisation de l'intelligence artificielle comporte d'énormes avantages et opportunités, mais présente également des risques très importants. Le principal, et qui est déjà vérifié, est le risque de dissimulation ou d'intensification de la discrimination. Dans l'attente d'une éventuelle réglementation européenne - annoncée après l'avoïr élaboré-, le présent travail s'intéresse aux causes de cette discrimination, ainsi que les problèmes qu'elle pose aux normatif anti-discrimination et l'important rôle de la réglementation européenne en matière de protection des données. L'étude se termine en soulignant la nécessité d'une intervention multidimensionnelle, que non seulement évite les éventuelles inégalités et exclusions sociales motivées par l'application de l'IA, mais garantit également une IA sûre, éthique et robuste, avec le soutien de la responsabilité sociale des entreprises.

*Mots-clés:* algorithme, discrimination, intelligence artificielle.

SUMARIO: I. *Introducción: elogio y peligros del Big Data.* II. *Algoritmo y prohibición de discriminación.* III. *Algoritmo y Protección de datos.* IV. *La necesidad de una visión más amplia.* VI. *Y de una IA éticamente responsable: el papel de la Responsabilidad Social.* VI. *Conclusiones.* VII. *Bibliografía*

## I. INTRODUCCIÓN: ELOGIO Y PELIGROS DEL BIG DATA

**E**l imparable desarrollo de las TIC y su rápida y masiva incorporación a todos y cada uno de los ámbitos de la actividad humana no sólo están provocando una nueva y profunda revolución “industrial”, sino también una

no menos intensa transformación de las formas de comportamiento personal y profesional que impactan en casi todos los ámbitos de nuestra realidad. El desarrollo de nuevas técnicas de computación, la generalización del Internet de las personas y de las cosas, junto con las nuevas posibilidades de tratamiento masivo de datos, están produciendo cambios económicos, políticos y sociales que afectan, de forma progresivamente acelerada, desde la propia concepción de la empresa y del trabajo, a las relaciones sociales e interpersonales, sin olvidar su notable impacto en los ámbitos culturales y políticos, en una aldea seguramente mucho más global de lo que en los años sesenta pudiera haber llegado a imaginar McLuhan.

Pues bien, uno de los vectores más importantes en todo este proceso de transformación económica y social ha sido, y sin duda es, la progresiva importancia atribuida, de forma explícita o implícita, a las propuestas o predicciones que realizan las propias aplicaciones y programas en nuestra toma de decisiones diarias. Es un hecho constatable que buena parte de nosotros ya nos hemos acostumbrado, e incluso agradecemos, a que estas aplicaciones —que, además, nos siguen constantemente a través de dispositivos permanentemente portados (“*wareables*”)— nos “sugieran” u “orienten” en todas y cada una de nuestras elecciones y decisiones: desde el tiempo y el camino a seguir para llegar a nuestro destino, a los restaurantes o tiendas a las que ir, la ropa o los zapatos que comprar, los libros que adquirir o leer (*electronic butler*), pasando por las propuestas de qué películas ver o qué tipo de música, según nuestras elecciones previas, oír. Y todo ello sin olvidar, lo que a mi juicio resulta aún más preocupante, con qué personas contactar y, sobre todo, qué noticias leer o visualizar en nuestras cada vez más segmentadas —pero, al parecer, importantes— redes sociales. Y si esto es así en el ámbito de las decisiones personales, aún resulta más claro e intenso, por razones obvias, en el caso de las decisiones de trascendencia colectiva que tanto las administraciones, como sobre todo las empresas, deben adoptar en su actividad diaria.

La novedosa capacidad de almacenar, clasificar y tratar de forma automática en busca de patrones (*data mining*), cantidades ingentes de datos (economía de los datos<sup>1</sup>), proporcionados de forma aparentemente voluntaria —en un sin-

---

<sup>1</sup> Sobre este “ecosistema” que permite a diversos agentes del mercado extraer valor de una amplia gama de datos que deben ser accesible y utilizables, véase, por todos, la Comunicación de la Comisión Europea, *La Construcción de una economía de los datos europea*, Bruselas, 10.1.2017, COM(2017) 9 final, así como el Reglamento (UE) 2018/1807 del Parlamento Europeo y del Consejo, del 14 de noviembre de 2018, relativo a un marco para la libre circulación de datos no personales en la Unión Europea.

gular proceso de publicitación, propio también de nuestro tiempo— por cada vez más personas; unida, en segundo lugar, a las posibilidades de tratamiento aparentemente científico de los mismos por parte de unos programas que, además, pueden aprender por sí mismos (*learning machine*) en un constante proceso de prueba y error, han llevado a algunos a postular toda una nueva filosofía que podía calificarse como “dataísmo”, y que auguraba, al menos aparentemente, una segunda Ilustración (Byung-Chul Han);<sup>2</sup> una nueva era de la razón, libre de prejuicios y meras intuiciones y en el que el dato, y las ciencias y procesos de la Inteligencia Artificial (IA)<sup>3</sup> permitirían hallar nuevas correlaciones —obsérvese, no necesariamente causalidades—<sup>4</sup> que permitiría significativas ganancias en eficiencia y objetividad, evitando discriminaciones y sesgos propios de los humanos, mejorando no sólo nuestras decisiones privadas, sino también la equidad, la justicia y, sobre todo, la eficiencia de los agentes económicos y sociales, tanto públicos como privados.

En palabras del Parlamento Europeo, en su Resolución del 14 de marzo de 2017,<sup>5</sup> esta “analítica” presentaría un indudable potencial para ciudadanos, académicos, la comunidad científica y los sectores público y privados, al generar, no sólo importantes oportunidades para los ciudadanos, sino también al contribuir a la mejora de “la optimización y eficiencia de las empresas”. Estos instrumentos podrían así “ofrecer una ventaja competitiva para los procesos de toma de decisiones de las empresas europeas, al mismo tiempo que “el sector público” podría igualmente “beneficiarse de una mayor eficacia gracias a un mejor conocimiento de los distintos niveles de desarrollo socioeconómico”.

En este contexto, no debe extrañarnos que una buena parte de los países industrializados, y con ellos, obviamente, la Unión Europea,<sup>6</sup> hayan lanzado du-

---

<sup>2</sup> Para una perspectiva crítica, que aquí se comparte plenamente, Han, Byung-Chul, *Psicopolítica. Neoliberalismo y nuevas técnicas de poder*, trad. de Alfredo Bergés, Barcelona, Herder, 2014, pp. 87 y ss.

<sup>3</sup> Ante la multiplicidad de definiciones utilizadas, en este trabajo vamos a seguir las establecidas por el Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial, *Una definición de la inteligencia artificial: Principales capacidades y disciplinas científicas*, Bruselas, Comisión Europea, abril de 2019.

<sup>4</sup> Y ello ya que, como con extraordinario acierto sintetizaba Han, Byung-Chul, *Psicopolítica. Neoliberalismo...*, cit., p. 102, “las correlaciones sustituyen a las causalidades. El así sustituye al por qué”.

<sup>5</sup> Procedimiento 2016/2225(INI).

<sup>6</sup> Para un repaso de las actuaciones en este campo de la UE —cuyo inicio puede situarse seguramente en la Comunicación de la Comisión Europea, *Inteligencia artificial para Europa*, Bruselas 25.4.2018, COM(2018)237 final— nos remitimos, por brevedad, a lo que señala la Comu-

rante estos últimos años una amplia batería de estrategias y actuaciones en este campo, entre las que debe situarse, por ejemplo, la reciente Estrategia Española de I+D+I en Inteligencia Artificial.<sup>7</sup> La IA se ha convertido así en uno de los ejes de actuación de la política mundial y de la atención tanto de empresas como de organizaciones internacionales. Y ello ya que, como se ha señalado recientemente, la IA podría añadir alrededor de 14 billones (millones de millones) de euros a la economía global en el año 2030 y duplicar las tasas de crecimiento económico para 2035, al mismo tiempo que proporcionaría una amplia gama de oportunidades para las empresas al permitirles “una mejor comprensión de sus procesos, las necesidades de sus clientes y socios, y el entorno empresarial en general”.<sup>8</sup> De ahí que tampoco deba sorprendernos cómo la financiación global de capital de riesgo para nuevas empresas que se especializan en aplicaciones de inteligencia artificial haya tenido una tasa de crecimiento anual del 85% entre 2012 y 2017, o que la financiación se triplicase con creces entre 2016 y 2017, para alcanzar más de 11 000 millones de euros.<sup>9</sup>

El resultado de todo ello es, como recordaba hace ya algunos años Cofone, que estas decisiones, basadas o incluso automáticamente generadas por instrumentos de inteligencia artificial, nos afectan prácticamente a todos, de forma frecuente, y en ocasiones incluso sin nuestro conocimiento: ya sea como consumidor de productos financieros o de seguros de salud, como público objetivo de específicas campañas publicitarias, como simple aspirante a un puesto de trabajo o como empleado sometido a procesos de promoción o de evaluación del desempeño.<sup>10</sup>

---

nicación de la Comisión Europea, *Fostering a European approach to Artificial Intelligence* Bruselas, 21.4.2021 COM(2021) 205 final.

<sup>7</sup> Ministerio de Ciencia, Innovación y Universidades, 2019, [https://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia\\_Inteligencia\\_Artificial\\_IDI.pdf](https://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_IDI.pdf). Una recapitulación de las estrategias de los distintos países, a la que por brevedad nos remitimos, pp. 12 y 13 de este mismo documento.

<sup>8</sup> En este sentido, y por mencionar sólo un ejemplo, recientemente se ha señalado —Fernández, Ana, “Inteligencia artificial en los servicios financieros”, *Boletín Económico Banco de España*, núm. 2, 2019, p. 2— cómo “la aplicación de técnicas de inteligencia artificial en la prestación de servicios financieros puede dar lugar a mejoras de eficiencia, reducción de costes, incrementos de calidad, aumento del nivel de satisfacción de los clientes o mayor inclusión financiera”.

<sup>9</sup> En un sentido similar, European Political Strategy Centre, *The Age of Artificial Intelligence*, EPSC Strategic Notes, núm. 29, 2018.

<sup>10</sup> Cofone, Ignacio, “Algorithmic Discrimination Is an Information Problem”, *Hastings Law Journal*, vol. 70, 2019, pp. 1391 y ss.

El problema, como siempre, es que esta, como cualquier otra innovación técnica que se aplica rápidamente al sistema productivo y al entorno social, puede tener una doble vertiente; un carácter bifronte,<sup>11</sup> especialmente preocupante ante la nueva generación de algoritmos basados en su propio aprendizaje.<sup>12</sup>

Sin entrar aquí en cuestiones de una extraordinaria importancia como el uso de estas herramientas de forma autónoma en combate,<sup>13</sup> o sus posibles efectos de segmentación social y radicalización política —al hacernos leer y, por tanto creer, solo en opiniones cercanas a la nuestra, desconociendo e invisibilizando a “los otros”—, lo que sí parece evidente es que estas mismas ciencias y técnicas de IA pueden convertirse también en un elemento extraordinariamente peligroso en relación, especialmente, con el disfrute y ejercicio de los derechos humanos y, singularmente, con la prohibición de discriminación;<sup>14</sup> ya sea esta entendida como un derecho al igual disfrute de los derechos en relación con ciertos rasgos, o como elemento de igualdad, no sólo formal sino también material, de grupos minoritarios tradicionalmente segregados o alejados del proceso de poder y sometidos a los consiguientes prejuicios injustificados.<sup>15</sup>

---

<sup>11</sup> Véase, ampliamente, Comunicación de la Comisión Europea, *Fostering a European...*, cit., pp. 1 y ss. Al otro lado del Atlántico, esta posibilidad ya fue puesta de manifiesto en 2014 por el conocido Informe Podesta —Executive Office of the President, *Big Data: Seizing Opportunities, Preserving Values*, mayo de 2014, pp. 64 y ss.— y desarrollada en relación con diversos campos (acceso al crédito, empleo, educación superior, justicia penal) por el Informe de 2016 —Executive Office of the President, *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*, mayo de 2016—. En la doctrina, por todos, Sancho Villa, Diana, “Las decisiones individuales automatizadas, incluida la elaboración de perfiles”, en Troncoso, Antonio (dir.), *Comentario al Reglamento General de Protección de Datos y a la Ley Orgánica de Protección de datos personales y garantía de los derechos digitales*, Pamplona-Navarra, Civitas-Thomson Reuters, 2021, t. I, p. 1729.

<sup>12</sup> Tutt, Andrew, “An FDA for Algorithms”, *Administrative Law Review*, vol. 69, núm.1, invierno de 2017. Esta dualidad seguramente fue descrita elocuentemente por Stephen Hawking cuando, en una de sus últimas intervenciones, en la Web Summit de 2017, señalaba cómo “el éxito en la creación de una IA eficaz podría ser el evento más importante en la historia de nuestra civilización. O el peor. Simplemente no lo sabemos. Y ello ya que no podemos saber si la IA nos ayudará infinitamente, si nos ignorará y nos dejará de lado, o posiblemente nos destruirá”.

<sup>13</sup> Ampliamente, sobre este tema, véase Asaro, Peter, “What is an ‘AI Arms Race’ Anyway?”, *I/S: A Journal of Law for the Information Society*, vol. 15, núm. 1-2, primavera de 2019, pp. 45-64.

<sup>14</sup> Expresivamente señala Todolí, Adrián, “En cumplimiento de la primera Ley de la robótica: Análisis de los riesgos laborales asociados a un algoritmo/inteligencia artificial dirigiendo el trabajo”, *Labour & Law Issues*, vol. 5, núm. 2, 2019, p. 17 cómo “el procesamiento automatizado de datos incrementa exponencialmente las posibilidades de vulneración de los derechos de los trabajadores y sufrir discriminación”.

<sup>15</sup> Sobre las distintas percepciones del concepto de discriminación, véase, por todos, Barbera, Marzia, *Discriminazioni ed eguaglianza nel rapporto di lavoro*, Milán, Giuffrè, 1991, pp. 85 y ss.

Seguramente el ejemplo más destacado y conocido —aunque no el único—<sup>16</sup> de estos temores se centra en la conocida aplicación COMPAS —*Correctional Offender Management Profiling for Alternative Sanctions*—, un software de evaluación del riesgo de posible reincidencia y peligrosidad, propiedad inicialmente de la empresa Northpointe, y aplicado fundamentalmente en los procesos penales en Estados Unidos. Más allá el éxito o no de sus predicciones y del progresivo impacto que las mismas han tenido en los procesos penales estadounidenses, lo llamativo es que según ciertos estudios<sup>17</sup> —eso sí, criticados metodológicamente por la propia empresa propietaria—<sup>18</sup> el programa mostraría un claro sesgo frente a los afroamericanos. Y así —siempre según aquel estudio—, mientras sólo un 28 % de los afroamericanos que fue etiquetado de menor riesgo reincidió, esta misma cifra se elevaba al 47.7 % en el caso de los caucásicos. Y este mismo sesgo se reflejaría —siempre según el estudio de *ProPublica*— en el hecho de que mientras sólo el 23.5 % de los blancos etiquetados como de mayor riesgo no reincidió, esta misma cifra se incrementaba hasta el 44.9 % en el caso de la población de color.<sup>19</sup> Y este caso no ha sido desgraciadamente el único. En marzo de 2018, la *National Fair Housing Alliance* (NFHA), junto con otras organizaciones, planteó una demanda frente a *Facebook* por permitir a los anunciantes discriminar —al no mostrarles sus anuncios— a grupos tradicionalmente postergados, en razón, entre otros, de su raza u origen nacional. Aunque en este caso no se llegó a juicio, lo cierto es que en el acuerdo que puso fin al procedimiento se establecieron diversas vías para eliminar esta situación.<sup>20</sup> Y un sesgo similar, en razón del sexo, ha sido detectado, igualmente, y por señalar

<sup>16</sup> Otros ejemplos de efectos negativos, especialmente en relación con segmentos deprimidos (enfermos, personas sin hogar y menores en riesgo), en la conocida obra de Eubanks, Virginia, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, Nueva York, St. Martin's Press, 2017.

<sup>17</sup> Angwin, Julia *et al.*, “Machine Bias. There’s software used across the country to predict future criminals. And it’s biased against blacks”, *ProPublica*, 23 de mayo de 2016, disponible en: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

<sup>18</sup> Dieterich, William *et al.*, “COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity”, 8 de julio de 2016, disponible en: [http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica\\_Commentary\\_Final\\_070616.pdf](http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf).

<sup>19</sup> Debemos señalar, no obstante, que el 13 de julio de 2016 el Tribunal Supremo del Estado de Wisconsin dictó sentencia en el conocido caso *State v. Loomis*, admitiendo el posible uso de esta herramienta en la valoración de peligrosidad. Una mayor información sobre este tema en Martínez Garay, Lucía, “Peligrosidad, algoritmos y *due process*: el caso State vs. Loomis”, *Revista de Derecho Penal y Criminología*, núm. 20, 2018, pp. 485 y ss.

<sup>20</sup> Puede consultarse el mismo en <https://nationalfairhousing.org/wp-content/uploads/2019/03/FINAL-SIGNED-NFHA-FB-Settlement-Agreement-00368652x9CCC2.pdf>. Para un encuadr-

sólo algunos ejemplos más, en relación con los anuncios de empleo recibidos,<sup>21</sup> o en la programación utilizada por Amazon durante años para su contratación.<sup>22</sup>

De ahí que, a la espera de un posible marco general en la regulación de la Inteligencia Artificial, como el que muy recientemente ha planteado la Comisión Europea con su propuesta de Reglamento por el que se establecerían normas armonizadas sobre inteligencia artificial,<sup>23</sup> vayamos a abordar aquí algunas cuestiones que, desde la perspectiva exclusivamente antidiscriminatoria y de protección de datos, plantea la posible utilización de estos algoritmos y los sistemas de Inteligencia Artificial.

## II. ALGORITMO Y PROHIBICIÓN DE DISCRIMINACIÓN

Las razones por las que estas herramientas, que en un principio parecían llamadas a eliminar los sesgos, incluso inconscientes, de las decisiones humanas,<sup>24</sup> han podido acabar convirtiéndose, al menos en ocasiones, en una fuente de posibles discriminaciones, son, seguramente, varias.

En primer lugar, cabría pensar, al menos como hipotética posibilidad, en que la propia configuración del algoritmo esté articulada, por acción u omisión, consciente o incluso inconscientemente, como un instrumento que genere la discriminación. Y ello ya que el aura de objetividad y certeza “científica” que

---

amiento de esta cuestión véase Parker Black, Jacob, “Facebook and the Future of Fair Housing Online”, *Oklahoma Law Review*, Vol, 72, núm. 3, 2020, pp. 711 y ss.

<sup>21</sup> Véase, en relación con los anuncios de trabajo de Google, el ya clásico trabajo de Datta, Amit *et al.*, “Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination”, *Proceedings on Privacy Enhancing Technologies*, vol. 2015, núm. 1, pp. 92 y ss. Varios ejemplos de ello en Kim, Pauline T, “Data-Driven discrimination at work”, *William & Mary Law Review*, Vol. 58: 857, I. 3, 2017, pp. 863-864.

<sup>22</sup> Reuters, “Amazon scraps secret AI recruiting tool that showed bias against women”. 11 octubre 2018, <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>.

<sup>23</sup> La denominada Ley de Inteligencia artificial 2021/0106 (COD) aprobada por la Comisión una vez cerradas la versión inicial de estas líneas, el 21 de abril de 2021. El texto de la propuesta disponible en COM(2021) 206 final.

<sup>24</sup> Esta perspectiva, que parcialmente compartimos, sigue estando presente en autores como Kleinberg, Jon *et al.* “Discrimination in the Age of Algorithms”, *Journal of Legal Analysis*, Volume 10, 2018, pp. 113 y ss., para los que: “*With the appropriate requirements in place, algorithms create the potential for new forms of transparency and hence opportunities to detect discrimination that are otherwise unavailable... algorithms are not only a threat to be regulated; with the right safeguards, they can be a potential positive force for equity*”.

tradicionalmente se ha otorgado a estos instrumentos, unido a la frecuente opacidad del propio proceso interno de toma de decisiones —ya sea por la pretendida o real complejidad del algoritmo o por la imposibilidad de acceder al mismo, dada su protección como secreto o propiedad industrial— lo podrían convertir en un mecanismo especialmente opaco (la conocida *black box society* de la que hablaba hace ya años Pasquale<sup>25</sup>), destinado, como decimos, a “justificar” o, al menos a argumentar, decisiones objetivamente discriminatorias.<sup>26</sup>

A nuestro juicio, un ejemplo reciente de esta posibilidad podría encontrarse en el conocido algoritmo “Frank” de la empresa Deliveroo, al menos en su versión tradicional. Y ello porque, como señaló hace escasos meses la Sección Laboral del Tribunal de Bolonia (Italia) en su Resolución del 31 de diciembre de 2020,<sup>27</sup> este sistema —sustituido, eso sí, a finales de 2020— permitía, como tantos otros, reservar las fases horarias más interesantes para ciertos *riders* en función de parámetros de fiabilidad y participación en los picos de actividad. El problema en este caso era que el mismo no tenía en cuenta, por ejemplo, las ausencias motivadas por la adopción de medidas de conflicto colectivo. De ahí que al secundar la huelga y no aceptar encargos en fases previamente reservadas, el *riider* pudiera sufrir una drástica reducción de su puntuación y, por tanto, de sus futuras opciones de trabajo. Y de ahí, también, que para el Tribunal este algoritmo, al considerar irrelevante los motivos de la falta de participación, condujera a una situación —para la magistrada, voluntariamente asumida por la empresa— incardinable dentro del concepto de discriminación indirecta.<sup>28</sup>

En cualquier caso, y más allá de las dudas que esta última calificación pudiera provocar, y de que muchos de estos sesgos involuntarios puedan combatirse mediante procesos de supervisión que permitan analizar los requisitos, restricciones y decisiones del sistema de un modo claro y transparente,<sup>29</sup> lo que sí parece evidente es que este tipo de situaciones no parecen ser, no obstante, ni las más frecuentes, ni al menos a nuestro juicio, las más complejas o peligrosas. En

<sup>25</sup> Pasquale, Frank, *The Black Box Society. The Secret Algorithms that Control Money and Information*, Londres, Harvard University Press, 2015.

<sup>26</sup> Véase Barocas, Solon y Selbst, Andrew D., “Big Data’s Disparate Impact”, *California Law Review*, vol. 104, 2016, pp. 692 y ss.

<sup>27</sup> Disponible en: <https://files.lbr.cloud/public/2021-01/Italy%20Deliveroo%20algorithm%20decision.pdf?jeHeFXxuAMv.Eg6Nirvso92di.phSBul>.

<sup>28</sup> Sobre esta resolución pueden consultarse, por ejemplo, el análisis de Perulli, Adalberto, “La discriminazione algoritmica: brevi note introduttive a margine dell’Ordinanza del Tribunale di Bologna”, *LavoroDirittiEuropa Rivista nuova di Diritto del lavoro*, núm. 1, 2021.

<sup>29</sup> Como señalaba, nuevamente, el Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial, *Directrices éticas...*, cit., p. 23.

realidad, los problemas o dificultades en este campo parecen provenir de causas mucho más complejas y, desde luego, bastante más sibilinas.

En este sentido, no cabe desconocer, en primer lugar, que este posible tratamiento discriminatorio a través de las decisiones informatizadas puede ser resultado, no tanto de una decisión voluntariamente tomada por el o por los programadores, sino la consecuencia, incluso inconsciente,<sup>30</sup> de trasladar al código y al propio algoritmo los sesgos internos e incluso las opciones y valores propios de unos programadores que no presenten la deseable diversidad étnica, de género o cultural. De la misma manera que unos negociadores colectivos casi exclusivamente varones pueden sobrevalorar, incluso inconscientemente, ciertos rasgos tradicionalmente masculinos en la evaluación y, por tanto, retribución de puestos de trabajo, unos programadores mayoritariamente hombres caucásicos pueden trasladar a la construcción del flujo o al etiquetado de datos, valores y creencias básicamente de sus grupos, infravalorando los de los demás.<sup>31</sup> De ahí la importancia de que también entre los programadores y en los equipos de trabajo se incluya una necesaria diversidad étnica, racial, de género o religiosa que, unida a una previa reflexión del peso y uso de los posibles indicadores y del etiquetado correcto de los datos de entrenamiento prevea y anticipe el posible tratamiento peyorativo o el impacto desigual que el propio código podría tener entre los colectivos tradicionalmente discriminados.<sup>32</sup>

Pero, en tercer lugar, y este es sin duda el caso más peligroso, también es posible encontrarnos casos en los que, siendo el algoritmo en principio neutro respecto a estos factores discriminatorios, el resultado o incluso su propia evolución puede llegar a serlo como consecuencia, básicamente, de los datos que lo alimentan o de los que se nutre para alcanzar su propuesta o resolución.

Este sería el caso, por ejemplo, de aquellas aplicaciones, nuevamente muy usuales en las plataformas tecnológicas de transporte, cuyas elecciones o recomendaciones —básicas para la visualización del agente en la misma— se basan en las propias valoraciones que realizan los clientes. En estos casos, es dable

---

<sup>30</sup> En esta línea de *unintentional discrimination* véase, por ejemplo, el conocido trabajo de Barocas, Solon, Selbst, Andrew D., “Big Data’s...”, *cit.*, pp. 672 y ss.

<sup>31</sup> Sobre los posibles efectos discriminatorios de la delimitación de las variables utilizadas *-target variable-* y las clases asociadas *-associated class-* por los programadores véase Barocas, Solon, Selbst, Andrew D., “Big Data’s...”, *cit.*, pp. 677 y ss.

<sup>32</sup> Sobre esta necesidad, véase, por ejemplo, las propuestas en este sentido del Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, *Directrices éticas...*, *cit.*, pp. 23 y 29: “Lo ideal es que la diversidad no solo se materialice en los equipos en términos de género, cultura y edad, sino también de antecedentes profesionales y conjuntos de competencias”.

pensar que en un porcentaje, quizá no mayoritario, pero sí significativo en cuanto a sus posibles efectos, tales valoraciones podrían estar sesgadas, o incluso ser claramente discriminatorias por razón de raza, etnia o género —por señalar los rasgos más frecuentes—.<sup>33</sup> De la misma manera que cabría incluso pensar si estas plataformas pueden permanecer incólumes, aisladas como terceros, frente a posibles actos discriminatorios que en los “mercados” teóricamente creados por las mismas pudieran generar clientes y oferentes que concurren en los mismos.

Es este un campo extraordinariamente complejo tanto por la fragmentación que se produce en las posiciones discriminatorias, como por la diversidad de estas mismas empresas —plataformas horizontales y/o verticales— o, en su caso, prestadores de servicios de la sociedad de la información. Por tanto y sin entrar aquí en la discutida cuestión de si nos encontramos ante intermediarios de servicios —que podrían reclamar en su caso la inexistencia de una obligación general de supervisión ex artículo 15 de la Directiva 2000/31/CE del Parlamento Europeo y del Consejo, del 8 de junio de 2000— o, como sucede claramente en las plataformas verticales, auténticas empresas,<sup>34</sup> de lo que no cabe duda, es que, especialmente en este segundo caso, el efecto discriminatorio nunca podría justificarse por la empresa como la simple consecuencia de valoraciones o preferencias de los clientes cuando estas se encuentren claramente sesgadas por una causa discriminatoria.

Como hace tiempo señaló el Tribunal de Justicia de la Unión Europea en el conocido caso *Feryn*<sup>35</sup> una decisión discriminatoria, por ejemplo, en la contratación, no puede ni podría justificarse bajo ningún concepto por la actitud discriminatoria de sus clientes frente a sus empleados. Por tanto, y trasladando la cuestión a estos supuestos, la empresa no podría justificar la decisión señalando que el algoritmo de evaluación del desempeño, se basa en valoraciones de sus clientes cuando resulte estadísticamente demostrado que las mismas reflejan un claro sesgo frente a determinados colectivos por razón de raza o sexo, entre otros. En estos casos, creemos que la empresa, a través de su algoritmo,

---

<sup>33</sup> En relación con la discriminación por origen racial en las valoraciones en Uber o Lift, véase el ya clásico trabajo de Ge, Yanbo *et al.*, “Racial and gender discrimination in transportation network companies”, *NBER Working Paper Series*, Working Paper 22776, National Bureau of Economic Research, octubre de 2016, quienes ya entonces proponían la eliminación de nombres en la información o reserva de viaje.

<sup>34</sup> Como ya declaró el Tribunal de Justicia (Gran Sala) en su conocida Sentencia del 10 de abril de 2018, C-320/16 ECLI:EU:C:2018:221 en relación con Uber France SAS.

<sup>35</sup> Sentencia del Tribunal de Justicia (Sala Segunda) de 10 de julio de 2008, Asunto C-54/07, Recopilación de la Jurisprudencia 2008 I-05187; ECLI:EU:C:2008:397.

no puede simplemente aplicar y perpetuar situaciones discriminatorias: su obligación de no discriminar le debería llevar a establecer mecanismos que eviten esos sesgos —a través del uso de la misma IA, mediante la exclusión de aquellas valoraciones claramente discriminatorias, disonantes con la media, sin una justificación objetiva y que se centran, precisamente en tales colectivos— permitiendo una valoración libre de tales discriminaciones, aunque estas se encuentren en la propia sociedad. Y todo ello sin dejar de recordar la necesidad de que tales valoraciones, como datos personales del trabajador, se sometan a la normativa de protección de datos sobre la que más tarde volveremos, lo que supone la exigencia al responsable de que el tratamiento se desarrolle sobre datos exactos (artículo 5.1.d RGPS), debiendo informarse de ello al interesado (artículo 14 RGPD) que podría así reaccionar frente a este tipo de comportamientos.<sup>36</sup>

Sea como fuere, seguramente la hipótesis más problemática se centra en aquellos casos en los que es el propio proceso de *machine learning* o aprendizaje automático el que acaba provocando la toma de decisiones o recomendaciones sesgadas. Como ya señalamos al principio de estas reflexiones, esta rama de la inteligencia artificial se caracteriza porque son las propias “máquinas” las que “aprenden” mediante la búsqueda de patrones que detectarían, no ya tanto la tradicional causalidad humana, sino la simple correlación entre datos aparentemente alejados o absolutamente desconectados del resultado que se pretende predecir. En este sentido, este tipo de procesos no pretenden señalar una causalidad tradicional entre un rango de datos y el resultado pretendido, sino la simple conexión o estrecha correlación entre unos y otros.

El problema en estos casos puede proceder, a su vez, de al menos tres situaciones, en ocasiones, además, estrechamente interrelacionadas. En primer lugar, cabe la posibilidad de que sea la propia muestra que se ofrece al sistema la se encuentre sesgada,<sup>37</sup> básicamente por la falta de pluralidad y diversidad en sus datos.<sup>38</sup> Esto puede llevar, simplemente, al desconocimiento de las minorías y, por

---

<sup>36</sup> Más ampliamente Calvo Gallego, Fco. Javier “Reputación digital y efectos sobre los trabajadores: redes sociales para contratación, uso y límites”, en Rodríguez-Piñero, Miguel y Todolí, Adrián (dir.), *Vigilancia y control en el Derecho del Trabajo digital*, Navarra, Aranzadi, 2020, pp. 468 y sig.

<sup>37</sup> Y ello, ya que como con claridad señalaban Barocas, Solon, Selbst, Andrew D., “Big Data’s...”, *cit.*, pp. 680, “*biased training data leads to discriminatory models*”.

<sup>38</sup> Véase Buolamwini, Joy y Gebur, Timnit, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, vol. 81, 2018, disponible en: <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>. Igualmente, véase el tantas veces citado trabajo de Barocas, Solon y Selbst, Andrew D., “Big Data’s...”, *cit.*, pp. 684 y ss.; Hacker, Philipp, “Teaching Fairness to

tanto, a su exclusión, como ocurrió en el caso de las técnicas de reconocimiento facial cuyas bases mostraban sólo rostros caucásicos, con los consiguientes errores en mujeres subsaharianas o asiáticas.<sup>39</sup> La solución en este caso a un problema, por lo demás conocido, es, al menos en apariencia, simple: al igual que en los equipos, también las muestras deben ser analizadas buscando reflejar la diversidad social, con todo lo que ello supone, obsérvese, no sólo de eliminación de discriminaciones, sino también de mejora de la eficacia del sistema.<sup>40</sup>

En esta misma línea de muestras sesgadas, un segundo problema en relación con los datos que se introducen en el proceso, y del que se nutre la máquina, es el hecho de que en ocasiones la elaboración de patrones por el sistema de IA se hace con base en conjuntos de datos etiquetados —*Labeling examples*— que en muchas ocasiones reflejan las elecciones anteriores en la empresa, por ejemplo, de selección o promoción, o que incluyen datos desfasados y, por tanto, no adecuados a la realidad.<sup>41</sup> Y ello puede provocar que el algoritmo tienda a repetir o incluso a intensificar, ocultándolo además dentro de lógicas a veces no evidentes, este mismo sesgo tradicional,<sup>42</sup> por ejemplo, primando a hombres sobre mujeres, no con base en este rasgo, obviamente, sino con elementos que con mayor frecuencia cabe encontrar en uno u otro colectivo. En este caso, nuevamente, la solución pasa por un previo examen y “limpieza” de estas mismas bases de datos, que deben reflejar realidades actuales, y con decisiones y “etiquetados” que muestren la tendencia a la paridad y a la diversidad que, en definitiva, se busca en los resultados extraídos del sistema de IA.

Finalmente, esta última referencia a la valoración de posibles elementos más cercanos o estadísticamente más recurrentes en los colectivos tradicionalmente

---

Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination under EU Law”, *Common Market Law Review*, vol. 55, núm. 4, también disponible en: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3164973](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3164973), de donde tomamos la numeración en adelante, p. 6; Todolí, Adrián, “En cumplimiento...”, *cit.*, p. 16.

<sup>39</sup> Otro ejemplo bien conocido fue la utilización en Boston de los teléfonos móviles para detectar la existencia de baches. El problema es que los resultados estaban sesgados por la desigual distribución y presencia de teléfonos móviles entre las distintas clases sociales. Véase, nuevamente, Barocas, Solon y Selbst, Andrew D., “Big Data’s...”, *cit.*, p. 685.

<sup>40</sup> Véase, también en este sentido, Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, *Directrices éticas...*, *cit.*, p. 23.

<sup>41</sup> Con referencia a un conocido ejemplo, Hacker, Philipp “Teaching fairness...”, *cit.*, p. 6.

<sup>42</sup> Véase, nuevamente, Barocas, Solon y Selbst, Andrew D., “Big Data’s...”, *cit.*, pp. 681 y ss., quienes, además de señalar algunos ejemplos, sintetizaban la cuestión al indicar que “*as prior decisions affected by some form of prejudice serve as examples of correctly rendered determinations, data mining will necessarily infer rules that exhibit the same prejudice*”.

discriminados nos lleva, sin duda, al conocido y complejo problema de los proxies.<sup>43</sup> Y ello ya que el uso de bases cada vez más amplias y ligadas a aspectos sociales puede permitir la incorporación de datos que, sin ser los determinantes directos de la inclusión del sujeto en un grupo discriminado —color, sexo, etnia...— sí pueden permitir categorizarlos indirectamente —“*redundant encoding*” — en grupos que tiendan a coincidir con aquellos.<sup>44</sup>

En estos casos es importante diferenciar, al menos a nuestro juicio, el concepto de *proxy*, como elemento neutral, pero estadísticamente más frecuente en los colectivos tradicionalmente discriminados, de aquellos otros aspectos que, sin referirse directa y expresamente a uno de los grupos discriminados, sí guardan una relación tal de pertenencia o conexión con el grupo tradicionalmente discriminado que, al menos a nuestro juicio, debieran simplemente considerarse datos identificatorios de colectivos discriminados y, por tanto, eliminarse de la información suministrada a la IA.

Pero más allá de estos casos, es obvia la existencia de estos proxies. Piénsese, por ejemplo, en el caso de la tradicional segregación por razas o etnias en determinadas zonas geográficas y en la posible utilización de factores tan aparentemente neutrales como el código postal —por ejemplo, en relación con la proximidad al centro del trabajo—, o, en el caso estadounidense, la prevalencia o minusvaloración de ciertas universidades cuyos alumnos mayoritariamente pertenecen a determinadas razas o sexos. Es posible que, dada la tradicional postergación de estos colectivos discriminados, la máquina, en su aprendizaje autónomo, detecte una notable correlación de tal dato —aparentemente, y solo aparentemente neutro—, con un menor nivel de estudios, un mayor grado de delincuencia o un mayor índice de posibles impagos. La consecuencia, por tanto, de aplicar sin más tal correlación sería finalmente el encasillamiento del sujeto dentro de la caracterización general del grupo y, por tanto, la aplicación al mismo de una correlación general, sustancialmente conectada con factores prohibidos, con independencia de las circunstancias concretas del sujeto afectado, de la propia evolución social —con todo lo que ello supone de perpetuación de la situación discriminatoria que tradicionalmente ha padecido ese colectivo—, o incluso, obsérvese, de que el propio sujeto forme parte o no de tales colecti-

<sup>43</sup> Sobre el mismo, nuevamente, Barocas, Solon y Selbst, Andrew D., “Big Data’s...”, *cit.*, pp. 691 y ss.; Todolí, Adrián, “En cumplimiento...”, *cit.*, p. 16.

<sup>44</sup> Cuando la utilización de estos proxies se hiciera voluntariamente estaríamos seguramente ante lo que Barocas, Solon, “Data mining and the discourse on discrimination”, *Proceedings of the Data Ethics Workshop*, Nueva York, 2015, denominaba “Inferring Membership in Protected Class”,

vos, lo que, evidentemente, no impediría la posible existencia de un fenómeno de discriminación.

En definitiva, frente a todas estas situaciones resulta evidente la posibilidad de utilizar el aparato represivo antidiscriminatorio no sólo nacional, sino también internacional y comunitario. En estos casos, la utilización de una IA —incluso aunque ésta o su software no sean propiedad de la propia organización que la utiliza— no priva de responsabilidad a la misma organización que adopta la decisión objetivamente discriminatoria. Y, de hecho, la obligación de garantizar un comportamiento no discriminatorio obliga incluso, *in eligendo* a sus proveedores o *in vigilando* su comportamiento, a que el resultado sea objetivamente no discriminatorio. Y ello con independencia de las evidentes mayores exigencias que, en especial en las IA cuyo uso implique un mayor riesgo, debieran imponerse al creador/proveedor de la misma.

El problema, por consiguiente, no es ya tanto la aplicabilidad de todo este cuerpo normativo —aunque ciertamente puedan subsistir algunas zonas de incertidumbre sobre todo en el ámbito de la más específica normativa europea o comunitaria—<sup>45</sup> sino la posible efectividad de tales normas en estos casos. Y ello, seguramente, por dos factores fundamentales.

En primer lugar, por la dificultad de detectar, sobre todo, las discriminaciones indirectas, es bien sabido cómo en muchos de estos casos la decisión afecta a individuos aislados que no tienen capacidad para conocer la existencia de posibles sesgos o impactos desiguales desde un punto de vista estadístico. Por ello, es extraordinariamente importante que, como veremos más adelante, la propia organización que utilice la IA o sus encargados realicen auditorías internas en las que se evalúen estos posibles efectos discriminatorios. Y, en segundo lugar, resulta absolutamente necesario —al menos a nuestro juicio— que en este ámbito los organismos de promoción de la igualdad de trato desarrollen, con el lógico apoyo y respaldo público, y los consiguientes medios, análisis de estas cuestiones sobre todo cuando se refieran a bienes o servicios de uso generalizado y/o necesidad social.<sup>46</sup> Y ello, lógicamente, estableciendo el correlativo deber de información de las empresas u organizaciones titulares de la IA.

La segunda dificultad se centra, lógicamente, en la oscuridad de este proceso. En primer lugar, por la propia oscuridad del proceso de toma de decisiones por la IA, especialmente cuando este es autónomo, lo que hace muy difícil o en

<sup>45</sup> Sobre el tema, resulta de interés Hacker, Philipp, “Teaching fairness...”, pp. 12 y ss.

<sup>46</sup> En esta línea, por ejemplo, ya se movía el conocido Informe Podesta - Executive Office of the President, *Big data: seizing..., cit.*, p. 65.

algunas ocasiones prácticamente imposible acceder al mismo por parte incluso de sus propios creadores. Y, en segundo lugar, por las mismas dificultades de acceder al propio algoritmo cuando este es la clave de la propia empresa y, por tanto, está protegida por el secreto empresarial.<sup>47</sup>

En estos casos, y más allá de la regla general, quizá sea importante recordar, no solo las reglas generales de prueba en estos casos, sino también la lógica subyacente en la doctrina del Tribunal de Justicia en estos casos. Como ya se señaló hace muchos años en el marco de la discriminación salarial de la mujer, la oscuridad de un proceso debe ser un argumento que traslade a la otra parte la prueba de su suficiente justificación cuando existan datos estadísticos suficientes.<sup>48</sup>

Pero también es cierto, que aun cuando el Tribunal haya dado pasos en este sentido, no ha avanzado lo suficiente en relación con el derecho de un candidato objetivamente apto a conocer las causas de su rechazo. Tanto en los asuntos *Kelly*<sup>49</sup> como *Meister*<sup>50</sup> el Tribunal de Justicia, refiriéndose, eso sí, a la normativa discriminatoria, señaló expresamente que tales normas “no prevén el derecho de un trabajador que alega de forma verosímil que reúne las condiciones enunciadas en un anuncio de contratación” a “acceder a la información que indica si al término del proceso de selección el empresario ha contratado a otro candidato”.

Es cierto que en ambos casos esta premisa se matizaba al señalarse que no cabría “excluir que una denegación total de acceso a la información por una parte demandada” pudiera “constituir uno de los factores que se deben tener en cuenta en el contexto de la acreditación de los hechos que permiten presu-

---

<sup>47</sup> En este sentido, si bien es cierto que la Directiva 2009/24/CE del 23 de abril de 2009, sobre la protección jurídica de programas de ordenador, en su Considerando 11, señala que “debe establecerse claramente que solo se protege la expresión del programa de ordenador y que las ideas y principios implícitos en los elementos del programa, incluidas las de sus interfaces, no pueden acogerse a la protección de los derechos de autor con arreglo a la presente Directiva”, no es menos cierto que seguramente dicho algoritmo sí puede considerarse tutelado por la Directiva (UE) 2016/943 del Parlamento Europeo y del Consejo, del 8 de junio de 2016, relativa a la protección de los conocimientos técnicos y la información empresarial no divulgados (secretos comerciales) contra su obtención, utilización y revelación ilícitas siempre que reúna los requisitos del artículo 2; esto es, que sea secreta, tenga un valor comercial y haya sido objeto de medidas razonables para mantenerla secreta.

<sup>48</sup> STJ de 17 de octubre de 1989, asunto 109/88, Danfoss, Recopilación de Jurisprudencia, 1989 03199, ECLI:EU:C:1989:383. En esta línea véase Hacker, Philipp, “Teaching fairness...”, p. 23.

<sup>49</sup> STJ (Sala Segunda) de 21 de julio de 2011, Asunto C-104/10, Kelly, Recopilación de Jurisprudencia 2011 I-06813; ECLI:EU:C:2011:506.

<sup>50</sup> STJ (Sala Segunda) de 19 de abril de 2012 asunto C-415/10, Meister, ECLI:EU:C:2012:217.

mir la existencia de una discriminación directa o indirecta”. Pero esta afirmación se hacía sólo como elemento que debería verificar el correspondiente tribunal nacional, tomando en consideración todas las circunstancias del litigio del que conoce. Y la práctica parece haber avanzado poco en este ámbito.

En cualquier caso, lo que también resulta evidente es el relevante papel que en este ámbito debieran jugar, nuevamente, los organismos de lucha contra la discriminación, tanto en el acceso como en la propia evaluación de estos algoritmos y resultados. Cuestión distinta es que, más allá de la necesaria regulación legal en este sentido, sean precisamente estos organismos los técnicamente más indicados para realizar esta cuestión.

### III. ALGORITMO Y PROTECCIÓN DE DATOS

Estas últimas reflexiones y, por qué no decirlo, carencias de la normativa antidiscriminatoria, nos conducen directamente al segundo cuerpo normativo igualmente implicado en estos temas y que, cuenta igualmente, con una amplia regulación supranacional. Nos referimos, claro está, al derecho fundamental a la protección de datos, reconocido no sólo por el artículo 18.4 de la Constitución Española, sino también por los artículos 8o. tanto del Convenio Europeo para la Protección de los Derechos Humanos y de las Libertades Fundamentales del 4 de noviembre de 1950 como de la Carta Comunitaria de los Derechos Fundamentales la Unión Europea. Se trata, además, de un derecho que en la actualidad cuenta con una regulación no sólo deducida del Convenio 108 del Consejo de Europa, del 28 de enero de 1981, sino también, y sobre todo, del Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, del 27 de abril de 2016, conectado, en el caso español, con la posterior Ley Orgánica 3/2018 del 5 de diciembre, de Protección de Datos Personales y Garantía de los Derechos Digitales (LOPDGDD) que, como veremos, puede resultar ciertamente de utilidad en este ámbito.<sup>51</sup>

En este sentido, y en relación con los posibles efectos discriminatorios de los instrumentos y sistemas de IA, lo primero que debe destacarse es que el juego conjunto del artículo 9o. del RGPD y de la LOPDGDD prohíbe cualquier tipo de tratamiento de datos personales que “revelen” —obsérvese, por tanto, que también, al menos a nuestro juicio, aquellos falsos proxies que en realidad identifican por un rasgo esta misma inclusión— el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, la afiliación sindical, la vida

<sup>51</sup> Véase, por ejemplo, Hacker, Philipp, “Teaching Fairness...”, pp. 24 y ss.

sexual o la orientación sexual de una persona física. Ello supone una importante limitación para su uso por la IA en la medida en que, en el caso concreto español, esta prohibición no puede ser “levantada” mediante el simple recurso de la autorización del interesado, admitiéndose por tanto tales tratamientos —estén estos automatizados o no— en los estrechos ámbitos que permiten las restantes letras del artículo 9.2 RGPD.

En cualquier caso, lo segundo que debe destacarse es la importancia que en este ámbito tiene la aplicación a tales procesos —como tratamiento que son, al menos en la amplísima definición del artículo 4.2 RGPD— de los principios de tratamiento establecidos en el artículo 5 RGPD —en especial los de licitud, lealtad, transparencia, limitación de datos, exactitud y responsabilidad proactiva—. Y ello ya que, como recuerda el Considerando 71 del RGPD, con el “fin de garantizar un tratamiento leal y transparente respecto del interesado”, el responsable del tratamiento ha de “utilizar procedimientos matemáticos o estadísticos adecuados para la elaboración de perfiles, aplicar medidas técnicas y organizativas apropiadas para garantizar, en particular, que se corrigen los factores que introducen inexactitudes en los datos personales y se reduce al máximo el riesgo de error”, de forma tal que se

tengan en cuenta los posibles riesgos para los intereses y derechos del interesado y se impidan, entre otras cosas, efectos discriminatorios en las personas físicas por motivos de raza u origen étnico, opiniones políticas, religión o creencias, afiliación sindical, condición genética o estado de salud u orientación sexual, o que den lugar a medidas que produzcan tal efecto. Las decisiones automatizadas y la elaboración de perfiles sobre la base de categorías particulares de datos personales únicamente deben permitirse en condiciones específicas.

Desde esta perspectiva, el artículo 35 RGPD establece además que, cuando

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

pudiendo en su caso, no sólo solicitar la consulta previa a la autoridad de control a la que hace referencia el artículo 36 RGPD, sino solicitar igualmente el asesoramiento del delegado de protección de datos del artículo 39 RGPD.

En cualquier caso, y en este contexto, nos interesa además resaltar aquí tres cuestiones incorporadas en esta normativa y con un impacto específico en relación con estas decisiones automatizadas.

El primero, y aparentemente fundamental, es el derecho del interesado<sup>52</sup> —recogido en el artículo 22 RGPD— a no ser objeto de determinadas decisiones basadas únicamente en un tratamiento automatizado, incluyendo entre estos la elaboración de perfiles (artículo 4.4 RGPD). Es esta una norma que, al menos en una primera lectura, podría parecer muy incisiva y de una extraordinaria importancia. Pero, sin embargo, basta un análisis algo más detenido de la misma para matizar significativamente esta primera impresión, a pesar de sus notables avances sobre la regulación europea anterior (artículo 15 Directiva 95/46/CE del 24 de octubre de 1995).

En este sentido, lo primero que llama la atención es que la norma europea no contempla literalmente una prohibición de la elaboración automatizada de una propuesta o previsión sobre un sujeto, con independencia de que tal tratamiento deba respetar obviamente las reglas y principios establecidos en el RGPD que señalamos anteriormente.<sup>53</sup> En realidad, y si se lee con atención, el artículo 22.1 RGPD tan solo contempla el derecho —como algo posterior a ese proceso o incluso al resultado del mismo— a no ser objeto o soportar la aplicación de dicha decisión cuando esta, además de significativa para el interesado,<sup>54</sup> se encuentre basada únicamente —y esta es la palabra clave— en el tratamiento automatizado. Y decimos que es esta la palabra clave ya que si acudimos a las versiones inglesa —“*solely*”— o francesa —“*exclusivement*”— podemos concluir que necesariamente quedarían fuera de este derecho —teniendo por tanto que soportarla sin más— todas aquellas situaciones en las que la IA solo realice una previsión que guíe la posterior decisión de un ser humano o incluso cuando ela-

<sup>52</sup> Lo que plantea problemas cuando los datos objeto de tratamiento han sido anonimizados, Sancho Villa, Diana, “Las decisiones individuales...”, *cit.*, p. 1735.

<sup>53</sup> En este mismo sentido, Sancho Villa, Diana, “Las decisiones individuales...”, *cit.*, p. 1729. Véase, no obstante, lo señalado por la misma autora en pp. 1739 y 1740.

<sup>54</sup> El considerando 71 considera, por ejemplo, significativas decisiones “como la denegación automática de una solicitud de crédito en línea o los servicios de contratación en red en los que no medie intervención humana alguna”. Sobre el tema y su mayor o menor expansión incluso en función de la vulnerabilidad del sujeto, véase, de nuevo, Sancho Villa, Diana “Las decisiones individuales...”, *cit.*, pp. 1736 y 1737.

bore una propuesta que pueda ser modificada, alterada o inaplicada por el responsable. En todos estos casos, en los que la persona ejerce una influencia real en el proceso de decisión, el artículo 22.1 RGPS no sería aplicable, si bien creemos que del espíritu de la norma se deduce que dicha participación “humana” tiene que ser significativa y, por tanto, estar en condiciones de aceptar, modificar o rechazar la propuesta o previsión que elabore el programa, no bastando, por ejemplo, su simple firma en un listado extraído directamente de la IA.<sup>55</sup>

Ahora bien, no es sólo esta limitación lo que mitiga la importancia real de esta norma. Además, y en segundo lugar, es necesario resaltar cómo el propio artículo 22.2 RGDPD excluye de este derecho a aquellas decisiones que, no solo estén autorizadas por la normativa europea o estatal —aunque eso sí, impone que estas normas deberán fijar medidas adecuadas para salvaguardar los derechos, libertades e intereses legítimos del interesado—, sino que también admite las mismas cuando se basen en el consentimiento explícito del interesado o, lo que es aún más importante, cuando aún sin este último consentimiento, dicha decisión sea “necesaria para la celebración o la ejecución de un contrato entre el interesado y un responsable del tratamiento”.

Sin poder detenernos aquí en una exégesis más detenida de este precepto, lo que sí es evidente es la necesidad, por simple coherencia, y para evitar la absoluta devaluación del derecho reconocido en el primer punto de ese mismo artículo, de, en primer lugar, garantizar que el consentimiento del interesado cumpla todas y cada una de las exigencias —libre, específico, informado e inequívoco— del artículo 4.11 RGDPD, teniendo en cuenta las observaciones recogidas en el Considerando 43 de la misma norma.<sup>56</sup> Y, en segundo lugar, que cuando este se omite, dicha decisión, basada únicamente en el tratamiento automatizado, sea realmente necesaria —por ejemplo, por el volumen de candidaturas a

---

<sup>55</sup> Esta parece ser la opinión del entonces Grupo de Trabajo sobre protección de datos del art. 29, *Directrices sobre decisiones individuales automatizada y elaboración de perfiles a los efectos del Reglamento 2016/679*, de 3 de octubre de 2017, WP251rev.01, de 6 de febrero de 2018, p. 23. En la doctrina, dando cuenta del debate, Sancho Villa, Diana, “Las decisiones individuales...”, cit., pp. 1732-1735 y 1738-1739. En el campo laboral, véase, por ejemplo, Todolí, Adrián, “La gobernanza colectiva de la protección de datos en las relaciones laborales: Big data, creación de perfiles, decisiones empresariales automatizadas y los derechos colectivos”, *Revista de Derecho Social*, núm. 84, 2018, pp. 8 y 9.

<sup>56</sup> Lo que, por ejemplo, lleva a Todolí, Adrián, “La gobernanza colectiva...”, cit., p. 10 a sostener que “dado que, en el contrato de trabajo el consentimiento del trabajador por regla general no será válido —y a falta de una regulación interna—, solamente se podrá entender válida la decisión automatizada si es necesaria para la celebración o ejecución del contrato de trabajo”.

analizar— y no meramente conveniente o simplemente más cómoda para su celebración o ejecución.<sup>57</sup>

En estos últimos casos, es cierto que el artículo 22.4 RGPD señala, obviamente, que las mismas “no se basarán en las categorías especiales de datos personales contempladas en el artículo 9.1 RGPD”, al que antes ya hemos hecho referencia. Pero incluso en ese caso, obsérvese que se admite una excepción cuando se apliquen las excepciones del apartado 2, letra *a)* o *g)* del mencionado artículo 9o. RGPD y “se hayan tomado medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado”.

Por ello, quizá mucho más interesante, al menos a efectos prácticos, resulte la obligación, contemplada en el artículo 22.3 RGPD, de acuerdo con la cual en aquellos casos en los que esta decisión, exclusivamente de la IA, se justifique por el consentimiento explícito del interesado o por ser necesaria la celebración o ejecución de un contrato, el responsable deberá al menos permitir al interesado, el derecho a expresar su punto de vista, a obtener una “intervención humana” por parte del responsable y, obviamente, a impugnar la decisión.<sup>58</sup>

Ahora bien, parece evidente que para ejercer este derecho —conectado obviamente al de oposición— resulta imprescindible la necesidad de una previa información. Y es aquí donde se aprecia nuevamente la importancia —sobre todo ante las limitaciones del derecho de información que antes señalamos en relación con la normativa antidiscriminatoria— de los artículos 13.2. f) y 14.2.f) RGPD. Y es que en ellos se establece no sólo la obligación de información del responsable del tratamiento al interesado —en función de que los datos se hayan obtenido o no de aquel— de la existencia de las decisiones automatizadas, incluida la elaboración de perfiles, sino también —y esto es lo importante—, el derecho a acceder —en conexión con el artículo 15.1 h) RGPD— a una “información significativa”, tanto sobre la importancia y las consecuencias previstas de dicho tratamiento para el interesado, como, sobre todo, de “la lógica aplicada”; un concepto, este último, ciertamente impreciso, pero que, al menos en nuestra opinión, debe ser interpretado desde la premisa general establecida en el mismo precepto y que reclama que esta información sea “necesaria para garantizar un tratamiento de datos leal” y, queremos destacar nosotros, “transparente”.<sup>59</sup>

<sup>57</sup> Sobre el tema véase, más ampliamente, Sancho Villa, Diana, “Las decisiones individuales...”, *cit.*, pp. 1740 y ss. y Todolí, Adrián, “La gobernanza colectiva...”, *cit.*, p. 10.

<sup>58</sup> Sobre la interpretación de estas garantías, nos remitimos nuevamente a Todolí, Adrián, “La gobernanza colectiva...”, *cit.*, pp. 10 y ss.

<sup>59</sup> Sobre los problemas y límites que ello plantea en los algoritmos de aprendizaje auto-

En cualquier caso, es importante también resaltar que, como suele ser habitual en la normativa de protección de datos, estos derechos de información se atribuyen de forma individual a cada interesado, con la consiguiente dificultad que ello conlleva para la obtención de resultados prácticos por cada uno de estos concretos individuos. El carácter transversal de esta normativa, su finalidad general y no sectorial, ha provocado que la misma olvide normalmente el posible momento colectivo de estos interesados; un momento colectivo que, sin embargo, resulta evidente en ciertos ámbitos como es el campo laboral.<sup>60</sup>

De hecho, y a pesar de la oportunidad, en gran parte perdida, que en este ámbito supuso la LOPDGD, al menos en España, parece evidente la existencia de una clara tendencia al reconocimiento de ciertos derechos de acceso a la lógica del algoritmo o de la IA a favor de los representantes legales de los trabajadores cuando estos elaboren decisiones de trascendencia objetiva para los empleados o candidatos al empleo.

Así, y por señalar sólo dos ejemplos, nos limitaremos aquí a recordar cómo el artículo 2o. de lo que parece será la propuesta de un inminente (en el momento de cerrar estas líneas) Real decreto-ley para la laboralización de las personas dedicadas al reparto en el ámbito de plataformas digitales —fruto del acuerdo en la mesa de diálogo social—, incorpora una nueva letra d) del artículo 64 del Estatuto de los Trabajadores; un nuevo precepto este, por el que se reconocería a la representación legal de los trabajadores —y de forma traslaticia a los delegados sindicales ex artículo 10.3 de la Ley Orgánica 11/1985, del 2 de agosto— el derecho a “ser informado por la empresa de los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, el acceso y mantenimiento del empleo, incluida la elaboración de perfiles”.

Sin poder detenernos aquí en un análisis detenido del precepto, lo que sí nos interesa destacar es, en primer lugar, que parece lógico considerar que esta norma no viola la Directiva (UE) 2016/943 relativa a la protección de los secretos comerciales, ya que el mismo artículo 3o. de la mencionada Directiva considera lícita la obtención cuando esta se realice en “el ejercicio del derecho de los tra-

---

mático, Sancho Villa, Diana “Las decisiones individuales...”, *cit.*, p. 1743 y Todolí, Adrián, “La gobernanza colectiva...”, *cit.*, p. 11 y ss.

<sup>60</sup> Sobre el tema me remito a lo que señalé específicamente en el campo laboral en Calvo, Fco. Javier, “Derechos digitales en la negociación colectiva”, *Comentario al Reglamento General de Protección de Datos y a la Ley Orgánica de protección de datos personales y garantía de los derechos digitales*. Pamplona, Civitas, pp. 4225 y ss. Desde otra perspectiva puede consultarse Todolí, Adrián, “La gobernanza colectiva...”, *cit.*, p. 14 y ss.

bajadores y los representantes de los trabajadores a ser informados y consultados de conformidad con... el derecho o las prácticas nacionales”; y, en segundo lugar, a destacar que no nos encontramos, a pesar de la denominación de la norma, ante una regla solo aplicable a estas concretas plataformas: su futura e hipotética ubicación en una norma general como es el Estatuto de los Trabajadores conduciría a concluir señalando que esta regla sería aplicable a toda empresa con representación legal, con independencia de su sector de actividad, lo que resulta de una enorme trascendencia práctica.

Y en esta misma línea, aún más significativo es, al menos a nuestro juicio, el artículo 80.5 del recientísimo XXIV Convenio colectivo del sector de la banca.<sup>61</sup> Bajo el título de “Derecho ante la inteligencia artificial”, este precepto, tras recordar el tantas veces señalado carácter bifronte de “las nuevas herramientas basadas en algoritmos”<sup>62</sup> reconoce —con la clara influencia del artículo 22.1 RGPD— no sólo el derecho de las personas trabajadoras “a no ser objeto de decisiones basadas única y exclusivamente en variables automatizadas, salvo en aquellos supuestos previstos por la Ley”, sino también el derecho a la “no discriminación en relación con las decisiones y procesos, cuando ambos estén basados únicamente en algoritmos”, pudiendo solicitar, en estos supuestos, “el concurso e intervención de las personas designadas a tal efecto por la Empresa, en caso de discrepancia”.

En cualquier caso, sin duda lo más relevante a estos efectos es que este mismo precepto convencional establece expresamente que:

las Empresas informarán a la representación legal de los trabajadores sobre el uso de la analítica de datos o los sistemas de inteligencia artificial cuando los procesos de toma de decisiones en materia de recursos humanos y relaciones laborales se basen, exclusivamente en modelos digitales sin intervención humana. Dicha información, como mínimo, abarcará los datos que nutren los algoritmos, la lógica de funcionamiento y la evaluación de los resultados.<sup>63</sup>

---

<sup>61</sup> Resolución de 17 de marzo de 2021, de la Dirección General de Trabajo, «BOE» núm. 76, de 30 de marzo de 2021.

<sup>62</sup> Literalmente señala como “las nuevas herramientas basadas en algoritmos pueden aportar valor hacia una gestión más eficiente de las Empresas, ofreciendo mejoras en sus sistemas de gestión. Sin embargo, el desarrollo creciente de la aportación de la tecnología requiere de una implantación cuidadosa cuando se aplica en el ámbito de las personas”.

<sup>63</sup> Sobre la experiencia italiana nos remitimos al excelente trabajo de Domínguez Morales, Ana, “Representación colectiva y negociación de derechos de trabajadores en plataformas”, *Revista Latinoamericana de Derecho Social*, núm. 29, 2019, pp. 80 y 81.

Finalmente, no querríamos terminar este breve apartado sin destacar una última cuestión: la importancia y la capacidad que en este ámbito pueden jugar las autoridades nacionales de control independientes a las que hace referencia el capítulo VI del RGPD. Y ello, no solo por la amplitud de sus funciones — artículo 57 RGOD— sino también por lo incisivo de sus poderes —artículo 58 RGPD— y de su capacidad sancionadora. De ahí que fuese seguramente deseable que estas autoridades desarrollaran un papel especialmente intenso en este campo, incluso de forma conjunta —y aquí retomamos la cuestión que abríamos al final del anterior apartado— con los organismos nacionales de lucha contra la discriminación.<sup>64</sup>

#### IV. LA NECESIDAD DE UNA VISIÓN MÁS AMPLIA

En cualquier caso, y en tercer lugar, resulta evidente que la importancia de la cuestión, su trascendencia no sólo jurídica, sino también económica, social y personal, exige que la lucha contra los posibles efectos discriminatorios de la IA no quede limitada únicamente al específico plano jurídico o normativo hasta ahora analizado o incluso otros sectoriales que aquí hemos obviado —como las normas de defensa de los consumidores o las importantes obligaciones en relación, sobre todo, para las plataformas en línea de muy gran tamaño, previstas en la Propuesta de Reglamento del Parlamento Europeo y del Consejo relativo a un mercado único de servicios digitales—.<sup>65</sup>

Sin rechazar la importancia de esta concreta faceta, sería injusto no mencionar al menos el impacto general, pero también social, que la denominada revolución algorítmica puede y seguramente tendrá, no sólo en el empleo o en la ocupación, sino también, y sobre todo, en la desigualdad y en la exclusión social; esto es, las posibles consecuencias segregadoras y discriminatorias derivadas de la implantación de la IA tanto en el plano de la desigualdad, como, y sobre todo,

---

<sup>64</sup> En esta línea, autores estadounidenses ya han propuesto la creación incluso de una agencia especializada en el control de estos algoritmos. Véase, por ejemplo, Tutt, Andrew, “An FDA...”, *cit.*, p. 116 y ss.

<sup>65</sup> La denominada “Ley de servicios digitales” COM (2020) 825, Véase en especial, en relación con los sistemas de recomendación, el artículo 29, de acuerdo con el cual estas singulares plataformas “establecerán en sus condiciones, de manera clara, accesible y fácil de comprender, los parámetros principales utilizados en sus sistemas de recomendación, así como cualquier opción que puedan haber puesto a disposición de los destinatarios del servicio para modificar o influir en dichos parámetros principales, incluida al menos una opción que no se base en la elaboración de perfiles”.

sobre la exclusión social y la discriminación por —en la variada terminología al uso— situación económica,<sup>66</sup> “fortuna”<sup>67</sup> “patrimonio”<sup>68</sup> o, más recientemente, aporofobia.<sup>69</sup>

En este sentido, quizá no esté de más recordar cómo, más allá del discutido efecto de toda revolución industrial sobre el volumen de empleo, sobre si estamos realmente ante el fin del mismo o simplemente ante una nueva modificación de su composición —y ante qué tipo de modificación (por ejemplo, en función de su mayor o menor capacidad para ser sustituidos por este nuevo tipo de máquinas)—<sup>70</sup> lo que sí parece indudable es que en todos estos procesos, al menos inicialmente, lo que sí se ha generado es un incremento de la desigualdad; la aparición de ganadores y de perdedores, con la consiguiente y posible segregación de nuevos o no tan nuevos colectivos. Piense, en este sentido, y por poner algunos ejemplos, en las dramáticas modificaciones en las demandas de las organizaciones y en las capacidades requeridas a los trabajadores; en la importancia que puede suponer el impacto en aquellos segmentos de más edad —en los que mayor es la brecha digital— o en aquellos con tareas más repetitivas (normalmente los salarios más bajos, menor retribución y cualificación) y en donde seguramente se centrará este proceso de destrucción de empleo, máxime ante sus mayores dificultades para lograr transiciones virtuosas en los nuevos mercados cada vez más transicionales. Si a todo ello unimos el posible impacto de determinadas formas empresariales asentadas en plataformas, en muchas de las cuales el modelo de trabajo a tiempo completo e indefinido, con un salario suficiente para satisfacer las necesidades familiares es casi una entelequia, tendremos las claves para comprender cómo este necesario acercamiento multidimensional debe abordar igualmente los posibles efectos de segregación y exclusión de un nuevo colectivo que podríamos definir como precariado<sup>71</sup> o del colectivo cada vez más amplio y, en ocasiones, estancado, de lo que se ha

<sup>66</sup> Artículo 2o. de la Declaración Universal de Derechos Humanos; artículo 26 del Pacto Internacional de Derechos Civiles y Políticos; artículo 2o. del Pacto Internacional de Derechos Económicos, Sociales y Culturales.

<sup>67</sup> Artículo 14 del Convenio Europeo de Derechos Humanos.

<sup>68</sup> Artículo 21 de la Carta de los Derechos Fundamentales de la Unión Europea.

<sup>69</sup> Cortina, Adela, *Aporofobia, el rechazo del pobre*, Barcelona, Paidós, 2017.

<sup>70</sup> Para una revisión de diversos estudios sobre estos efectos, desde la perspectiva centrada en su impacto en el volumen, pero también en la calidad del empleo, véase UGT, *Impacto de la automatización en el empleo en España*, octubre 2018 disponible en: [https://www.ugt.es/sites/default/files/24-\\_maquetado\\_impacto\\_de\\_la\\_automatizacion\\_en\\_el\\_empleo\\_en\\_espana.pdf](https://www.ugt.es/sites/default/files/24-_maquetado_impacto_de_la_automatizacion_en_el_empleo_en_espana.pdf).

<sup>71</sup> Utilizó aquí, obviamente, el conocido y discutido término creado por Standing, Guy, *The Precariat. The New Dangerous Class*, Londres, Bloomsbury Academic, 2011.

dado en llamar trabajadores o —más descriptiva y estadísticamente— ocupados en hogares pobres.<sup>72</sup>

En este sentido, esta nueva perspectiva, mucho más profunda y general, reclamaría también actuaciones dentro de una estrategia más extensa y multidimensional, que aborde desde una perspectiva deseablemente supranacional, las distintas dimensiones que, incluso desde el plano antidiscriminatorio, plantea la IA. Una perspectiva, que, en definitiva, asuma que sólo una IA segura y fiable, robusta técnicamente, pero también legalmente regulada y éticamente condicionada en sus múltiples impactos y dimensiones, permite eludir significativamente las consecuencias discriminatorias, segregadoras y de exclusión social que su uso puede provocar para una amplia capa de la sociedad.<sup>73</sup>

Por ello, creemos que solo estrategias complejas, centradas en las personas y sus necesidades, pueden ser efectivas; estrategias que, junto a medidas legales y promocionales por parte de los poderes públicos deben contemplar igualmente actuaciones que potencien —en palabras de la Comisión Europea— “la capacidad de anticipación” y “la priorización de la inversión en las personas” como “la piedra angular de un planteamiento integrador y centrado en el ser humano”, que eso sí, requerirá una inversión significativa, para lograr “que nadie quede rezagado”.<sup>74</sup>

#### V. Y DE UNA IA ÉTICAMENTE RESPONSABLE: EL PAPEL DE LA RESPONSABILIDAD SOCIAL

Sea como fuere, y ya para concluir estas breves notas, junto a ese desarrollo normativo y a la imprescindible actuación de los poderes públicos en la lucha contra los posibles efectos de segregación y exclusión social, creemos que esta estrategia omnicomprensiva debería igualmente incorporar un plano ético y hacer un llamamiento a las empresas y organizaciones privadas para que, en el marco de su responsabilidad social, desarrollaran códigos de conducta y exigencias de comportamiento que fueran más allá de este, eso sí imprescindible, marco jurídico y preceptivo.

<sup>72</sup> Véase, por todos, Calvo, Javier y Gómez-Alvarez, M. Rosario, *Trabajadores pobres y pobreza en la ocupación. Concepto y evolución de la pobreza en la ocupación: el impacto de las últimas reformas legales*, Murcia, Laborum, 2017.

<sup>73</sup> Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, *Directrices éticas...*, cit., pp. 5 y ss.

<sup>74</sup> En este sentido, Comisión Europea, *Inteligencia artificial...*, cit., pp. 13 y ss.

En esta línea, seguramente sería de un enorme interés, como hace poco tiempo señalaba el Grupo Independiente de Alto Nivel sobre Inteligencia Artificial, el desarrollo de códigos o políticas internas en el marco que podrían proporcionar futuros sistemas de normalización;<sup>75</sup> la adopción de estándares que permitieran abordar y eliminar en su caso estos sesgos en sus algoritmos y datos;<sup>76</sup> la obtención de certificaciones que, de forma complementaria a la regulación, dotaran de confianza a los consumidores e interesados; la creación de paneles o consejos de expertos para la supervisión y el asesoramiento en su funcionamiento, en el diseño y en la aplicación de tales sistemas; la designación, en su caso, de una persona encargada de las cuestiones éticas que pudieran plantearse en el desarrollo, implementación y utilización de estas IA; el intercambio de buenas prácticas con otras organizaciones; la creación de canales de información, debate e implicación con las partes interesadas y con los propios trabajadores a través del diálogo social, así como la creación de grupos y equipos de trabajo que reflejen la diversidad operando con mecanismos seguros desde su construcción y por defecto.<sup>77</sup>

En definitiva, actuaciones suplementarias, repetimos, a una normativa y a una actuación de los poderes públicos absolutamente necesaria, pero que permitiera incorporar en la dinámica y cultura de la propia empresa la necesidad de una IA lícita, ética y robusta —ya que su debilidad también puede provocar daños involuntarios—, adecuándola a las singularidades de cada sector y entidad; una actuación, por tanto, que partiendo de principios éticos como los propuestos recientemente por el Grupo de Alto Nivel —esto es, el respeto a la autonomía humana, la prevención del daño, la equidad y la explicabilidad— los integrara en el proceso de realización y evaluación de la misma, incluyendo igualmente sistemas de rendición de cuentas.

---

<sup>75</sup> En este ámbito seguramente serán de interés las aportaciones que en un futuro podría generar el Comité Técnico de Normalización CTN 71 Tecnologías Habilitadoras Digitales (THD) creado por la Asociación Española de Normalización (UNE) con el apoyo de la Secretaría de Estado de Avance Digital. En el marco de la ISO son de especial interés las labores del grupo de trabajo WG3 del SC 42, en especial ISO/IEC AWI TR 24368 “Overview of ethical and societal concerns” y ISO/IEC DTR 24027 “Bias in AI systems and AI aided decision making”.

<sup>76</sup> Véase, por ejemplo, el Proyecto P7003 - Consideraciones de sesgo algorítmico del IEEE, o, más en general, The IEEE Global Initiative on Ethics of Autonomous and Intelligent System disponible en: <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>.

<sup>77</sup> Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, *Directrices éticas...*, cit., pp. 28 y 29.

Unas actuaciones que, en definitiva, trasladaran al ámbito privado los principios hoy recogidos, sólo para el ámbito público en el artículo 21 de la Proposición de Ley integral para la igualdad de trato y la no discriminación, presentada por el Grupo Parlamentario Socialista en el Congreso;<sup>78</sup> una propuesta ésta que, no sólo parte de una acción promocional por parte de las administraciones públicas destinada a promover el uso de una Inteligencia Artificial ética y confiable —siguiendo especialmente las recomendaciones de la Unión Europea en este sentido—, sino que, además, impondría —aunque solo para los algoritmos “involucrados” en toma de decisiones que se utilicen por las administraciones públicas— la necesidad de que estos tuvieran en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas, incluyendo su diseño y datos de entrenamiento y abordando su potencial impacto discriminatorio. Y todo ello, claro está, sin olvidar la priorización de la transparencia en su diseño e implementación, así como la interpretabilidad de las decisiones adoptadas.

## VI. CONCLUSIONES

Tras todo lo hasta ahora expuesto, bien podríamos concluir señalando que la incorporación de la inteligencia artificial a casi todas las facetas de la vida social y, sobre todo, económica, no parece ser ya una mera o simple opción, sino una auténtica realidad, imprescindible, además, si no queremos quedarnos atrás en esta nueva economía del dato, basada en la actual revolución algorítmica.

El problema es que, aunque ciertamente esta revolución —y el uso de la IA como emblema— suponen innumerables ventajas y oportunidades, también presenta muy importantes riesgos. El primero de ellos es la posible aparición de discriminaciones, ya sea en la propia programación, o como consecuencia del proceso interno de aprendizaje, como consecuencia, en este último caso, de las posibles valoraciones sesgadas de los usuarios o, entre otros del carácter sesgado de los datos, tanto por su composición —infrarrepresentación de minorías— como por la presencia en los mismos de actuaciones discriminatorias pasadas, o por el uso de proxies que permiten reiterar, mediante elementos conexos, la categorización de grupos tradicionalmente discriminados, aplicándoles sin más los prejuicios derivados de realidades en muchas ocasiones desfasadas.

En todos estos casos, la normativa antidiscriminatoria, desarrollada incluso a nivel supranacional, supone un primer instrumento destinado a evitar estas con-

---

<sup>78</sup> Proposición de Ley integral para la igualdad de trato y la no discriminación (122/000121), *BOCG. Congreso de los Diputados* Núm. B-146-1 de 29/01/2021.

secuencias indeseables. Pero es evidente que esta normativa, aunque importante, se desarrolló mayoritariamente en otro contexto y que, por tanto, presenta ciertos déficits o dificultades en su aplicación a estos casos; unos déficits que no sólo reclaman una labor más incisiva de los organismos públicos competentes en esta materia, sino también, la ayuda de una normativa más moderna y adecuada, como la derivada de la protección de datos.

En este sentido, los principios generales de tratamiento —no sólo de transparencia, licitud, lealtad y limitación de finalidades, sino también el de exactitud, cuando los datos reflejan realidades desfasadas—, unido a los derechos de información sobre la lógica de decisiones automatizadas limitadas, aunque sea de forma muy condicionada, cuando no existe participación humana, ofrecen, como decimos, una batería normativa ciertamente significativa, más aún si se combinan con las reglas antidiscriminatorias y la exigencia de una evaluación de impacto en estos casos. De ahí la importancia de una actitud activa de las autoridades de control independientes que, al menos a mi juicio, deberían colaborar activamente en este campo con los organismos públicos de lucha contra la discriminación antes mencionados.

En este sentido, también sería importante que, siempre que sea posible —y destacadamente en el campo laboral, cuando el interés no sea simplemente difuso, sino que se encuentre claramente representado— se dé un papel relevante a los representantes colectivos de los interesados, por ejemplo, a la representación legal de los trabajadores, lo que sin duda abriría el campo fértil —como demuestra la experiencia española o italiana—, a una regulación no sólo legal, sino también colectiva del algoritmo y su funcionamiento.

En cualquier caso, también creemos que, junto a esta perspectiva, estrictamente normativa y específica, la nueva realidad que se abre ante nosotros exige, por un lado, una visión más amplia que abarque igualmente actuaciones ante los posibles efectos segregadores y de exclusión social que la generalización de estas tecnologías puede tener —sobre todo para los segmentos de mayor edad y menor cualificación—, y que, por el otro, potencie actuaciones de responsabilidad social, absolutamente necesarias y que podrían orientarse en línea con las propuestas del Grupo Independiente de Expertos de Alto nivel sobre Inteligencia Artificial de la Unión Europea.

## VII. BIBLIOGRAFÍA

- ANGWIN, Julia *et al.*, “Machine Bias. There’s software used across the country to predict future criminals. And it’s biased against blacks”, *ProPublica*, 23 de mayo de 2016.
- ASARO, Peter, “What is an ‘AI Arms Race’ Anyway?”, *I/S: A Journal of Law for the Information Society*, vol. 15, núm. 1-2, primavera de 2019.
- BARBERA, Marzia, *Discriminazioni ed eguaglianza nel rapporto di lavoro*, Milán, Giuffrè, 1991.
- BAROCAS, Solon, “Data mining and the discourse on discrimination”, *Proceedings of the data ethics workshop*, Nueva York, 2015.
- BAROCAS, Solon y SELBST, Andrew D., “Big Data’s Disparate Impact”, *California Law Review*, vol. 104, 2016.
- BUOLAMWINI, Joy y GEBRU, Timnit, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, vol. 81, 2018.
- CASTELLS, Manuel, *La era de la información: economía, sociedad y cultura*, 3a. ed., España, Alianza Editorial, 2005.
- CALVO, Fco, Javier, “Nuevas tecnologías y nuevas formas de trabajo”, *Creatividad y sociedad: revista de la Asociación para la Creatividad*, España, núm. 26, noviembre de 2016.
- CALVO GALLEGO, Fco. Javier, “Reputación digital y efectos sobre los trabajadores: redes sociales para contratación, uso y límites”, en RODRÍGUEZ-PIÑERO, Miguel y TODOLÍ, Adrián (dir.), *Vigilancia y control en el Derecho del Trabajo digital*, Navarra, Aranzadi, 2020.
- CALVO GALLEGO, Fco. Javier, “Derechos digitales en la negociación colectiva”, *Comentario al Reglamento General de Protección de Datos y a la Ley Orgánica de protección de datos personales y garantía de los derechos digitales*, Pamplona, Civitas, 2021.
- CALVO, Javier y GÓMEZ-ÁLVAREZ, M. Rosario (dir.), *Trabajadores pobres y pobreza en la ocupación. Concepto y evolución de la pobreza en la ocupación: el impacto de las últimas reformas legales*, Murcia, Laborum, 2017.
- COFONE, Ignacio, “Algorithmic Discrimination Is an Information Problem”, *Hastings Law Journal*, vol. 70, núm. 1389, 2019.
- Comisión Europea, *La Construcción de una economía de los datos europea*, Bruselas, 10.1.2017, COM (2017) 9 final.

- Comisión Europea, *Inteligencia artificial para Europa*’, Bruselas 25.4.2018, COM (2018) 237 final.
- Comisión Europea, *Fostering a European approach to Artificial Intelligence* Bruselas, 21.4.2021 COM(2021) 205 final.
- CORTINA, Adela, *Aporofobia, el rechazo del pobre*, Barcelona, Paidós, 2017.
- DATTA, Amit et al., “Automated Experiments on ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination”, *Proceedings on Privacy Enhancing Technologies*, vol. 2015, núm. 1,
- DIETERICH, William *et al.*, “COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity”, 8 de julio 2016, disponible en: [http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica\\_Commentary\\_Final\\_070616.pdf](http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf).
- DOMÍNGUEZ MORALES, Ana, “Representación colectiva y negociación de derechos de trabajadores en plataformas”, *Revista Latinoamericana de Derecho Social*, núm. 29, 2019.
- EUBANKS, Virginia, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, Nueva York, St. Martin’s Press, 2017.
- European Political Strategy Centre, *The Age of Artificial Intelligence*, EPSC Strategic Notes, núm. 29, 2018.
- Executive Office of the President, *Big data: seizing opportunities, preserving values*, mayo de 2014.
- Executive Office of the President, *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*, mayo de 2016.
- FERNÁNDEZ, Ana, “Inteligencia artificial en los servicios financieros”, *Boletín Económico Banco de España*, núm. 2, 2019.
- GE, Yanbo *et al.*, “Racial and gender discrimination in transportation network companies”, *NBER Working Paper Series*, Working Paper 22776, National Bureau of Economic Research, octubre de 2016.
- Grupo de Trabajo sobre protección de datos del art. 29, *Directrices sobre decisiones individuales automatizada y elaboración de perfiles a los efectos del Reglamento 2016/679*, de 3 de octubre de 2017, WP251rev.01, de 6 de febrero de 2018.
- Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial, *Una definición de la inteligencia artificial: Principales capacidades y disciplinas científicas*, Comisión Europea, Bruselas, abril de 2019.
- HACKER, Philipp, “Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law”, *Common Market Law Review*, vol. 55, núm. 4.

- HAN, Byung-Chul, *Psicopolítica. Neoliberalismo y nuevas técnicas de poder*, trad. de Alfredo Bergés, Barcelona, Herder, 2014.
- KLEINBERG, Jon *et al.*, “Discrimination in the Age of Algorithms”, *Journal of Legal Analysis*, vol. 10, 2018.
- KIM, Pauline T., “Data-Driven Discrimination at Work”, *William & Mary Law Review*, vol. 58, núm. 3, 2017.
- MARTÍNEZ GARAY, Lucía, “Peligrosidad, algoritmos y *due process*: el caso State vs. Loomis”, *Revista de Derecho Penal y Criminología*, núm. 20, 2018.
- PARKER BLACK, Jacob, “Facebook and the Future of Fair Housing Online”, *Oklahoma Law Review*, Vol. 72, núm. 3, 2020.
- PAQUALE, Frank, *The Black Box Society. The Secret Algorithms that Control Money and Information*, Londres, Harvard University Press, 2015.
- PERULLI, Adalberto “La discriminazione algoritmica: brevi note introduttive a margine dell’Ordinanza del Tribunale di Bologna”, *LavoroDirittiEuropa Rivista nuova di Diritto del lavoro*, num. 1, 2021.
- RIFKIN, Jeremy, *El fin del trabajo. Nuevas tecnologías contra puestos de trabajo: el nacimiento de una nueva era*, trad. de Guillermo Sánchez Gallego, Barcelona, Paidós Ibérica, 2010.
- SANCHO VILLA, Diana, “Las decisiones individuales automatizadas, incluida la elaboración de perfiles”, en TRONCOSO, Antonio (dir.), *Comentario al Reglamento General de Protección de Datos y a la Ley Orgánica de protección de datos personales y garantía de los derechos digitales*, t. I, Navarra, Civitas, Thomson Reuters, 2021.
- STANDING, Guy, *The Precariat. The New Dangerous Class*, Londres, Bloomsbury Academic, 2011.
- TODOLÍ, Adrián, “La gobernanza colectiva de la protección de datos en las relaciones laborales: Big data, creación de perfiles, decisiones empresariales automatizadas y los derechos colectivos”, *Revista de Derecho Social*, núm. 84, 2018.
- TODOLÍ, Adrián, “En cumplimiento de la primera Ley de la robótica: Análisis de los riesgos laborales asociados a un algoritmo/inteligencia artificial dirigiendo el trabajo”, *Labour & Law Issues*, vol. 5, núm. 2, 2019.
- TUTT, Andrew, “An FDA for Algorithms”, *Administrative Law Review*, vol. 69, núm.1, invierno de 2017.
- UGT, *Impacto de la automatización en el empleo en España*, octubre de 2018, disponible en: [https://www.ugt.es/sites/default/files/24-\\_maquetado-\\_impacto\\_de\\_la\\_automatizaci3n\\_en\\_el\\_empleo\\_en\\_espana.pdf](https://www.ugt.es/sites/default/files/24-_maquetado-_impacto_de_la_automatizaci3n_en_el_empleo_en_espana.pdf).