

# EL ALGORITMO COMO PROTAGONISTA DE LA RELACIÓN LABORAL. UN ANÁLISIS DESDE LA PERSPECTIVA DE LA PROHIBICIÓN DE DISCRIMINACIÓN\*

CARMEN SÁEZ LARA

*Catedrática de Derecho del Trabajo y de la Seguridad Social*

Universidad de Córdoba

## EXTRACTO

**Palabras clave:** Algoritmo, discriminación, legislación, ocupación y empleo

La inteligencia artificial se está desarrollando rápidamente y puede mejorar las condiciones de trabajo, actuando frente a las discriminaciones en las empresas. Aprovechar las oportunidades que ofrece la inteligencia artificial y abordar los retos que presenta, desde la perspectiva del derecho a la igualdad en las relaciones de trabajo, es el objeto de este trabajo, pues si los pronósticos se cumplen y la actual pandemia determina una aceleración digital, será preciso abordar la actualización de la tutela antidiscriminatoria, por parte de la normativa laboral. Para ello, se ha partido de identificar los rasgos que definen la discriminación algorítmica, como categoría jurídica, y frente a la cual la tutela antidiscriminatoria ha de actualizarse incorporando exigencias de transparencia, explicabilidad y evaluación o auditoría. Si el proceso de la IA es imparable, es preciso regularlo desde la protección de la igualdad y la promoción de la diversidad, también dentro de nuestras empresas, y por tanto promover el uso de algoritmos justos y equitativos.

## ABSTRACT

**Key words:** Algorithm, discrimination, legislation, occupation and employment

Artificial intelligence is developing rapidly and can improve working conditions, acting against discrimination in companies. Taking advantage of the opportunities offered by artificial intelligence and addressing the challenges it presents, from the perspective of the right to equality in labor relations, is the object of this work, because if the forecasts are met and the current pandemic determines a digital acceleration, it will be necessary to address the updating of anti-discrimination protection, by the labor regulations. For this purpose, the starting point has been to identify the features that define the algorithmic discrimination, as a legal category, and against which the anti-discrimination protection must be updated by incorporating transparency, explainability and evaluation or audit requirements. If the AI process is unstoppable, it must be regulated from the protection of equality and the promotion of diversity, also within our companies, and therefore promote the use of fair and equitable algorithms.

\* Este trabajo recoge la ponencia presentada al Congreso Internacional “Derechos Fundamentales, trabajo subordinado y digitalización” (octubre-noviembre de 2020), organizado por el sub-proyecto de investigación “Los derechos fundamentales ante el cambio del trabajo subordinado en la era digital” (DER2017-83488-C4-3-R), financiado por el Ministerio de Ciencia, Innovación y Universidades, la Agencia Estatal de Investigación, FEDER y por la Conselleria de Innovació, Universitats, Ciència y Societat Digital de la Generalitat Valenciana.

## ÍNDICE

1. LA ERA DE LOS ALGORITMOS Y LA RELACIÓN DE TRABAJO
2. DISCRIMINACIÓN ALGORÍTMICA: NUEVOS RASGOS Y NUEVOS RIESGOS
3. LA APLICACIÓN DE LA NORMATIVA ANTIDISCRIMINATORIA A LAS DECISIONES EMPRESARIALES AUTOMATIZADAS DISCRIMINATORIAS
4. PROPUESTA DE ACTUALIZACIÓN NORMATIVA EN ESTADOS UNIDOS
5. PROPUESTAS DESDE LA UNIÓN EUROPEA
  - 5.1. Estrategia Europea en IA: Marco ético respetuoso de derechos y valores
  - 5.2. El RGPD: garantía de las personas frente a las decisiones únicamente automatizadas
  - 5.3. El derecho de explicación sobre las decisiones basadas únicamente en procesos automatizados
  - 5.4. Propuestas de protección de los derechos fundamentales en el desarrollo y uso de algoritmos
6. A MODO DE RECAPITULACIÓN Y PROPUESTAS PARA ESPAÑA

## 1. LA ERA DE LOS ALGORITMOS Y LA RELACIÓN DE TRABAJO

Este siglo XXI se caracteriza por una importancia creciente de las decisiones automatizadas, de las decisiones basadas en aplicaciones de Inteligencia Artificial (IA), en sistemas algorítmicos<sup>1</sup>, que, con o sin intervención humana decisoria, producen efectos relevantes sobre las personas en el ámbito económico, social y jurídico. El análisis de macrodatos, a través de sistemas de aprendizaje automático, ha facilitado la creación de perfiles y ha automatizado las decisiones, en un número creciente de sectores, tanto privados como públicos<sup>2</sup>. El elevado volumen de información recopilada es precisamente el que requiere o hace necesarias formas de procesamiento automático, pues ningún cerebro humano podría tener en cuenta tal número de datos sobre un individuo. Igualmente en el mundo de la empresa, los algoritmos son ya herramientas de gestión empresarial. En 2016, Goldman Sachs anunciaba la introducción de un modelo algorítmico de gestión automática que incluía, entre otras decisiones de gestión de personal,

<sup>1</sup> Según el Grupo de Alto Nivel creado por la Comisión Europea, en el marco de su Estrategia sobre Inteligencia Artificial (IA), (Directrices éticas sobre una inteligencia artificial fiable, 2019) el término IA alude a los sistemas que manifiestan un comportamiento inteligente, al ser capaces de analizar el entorno y realizar acciones, con cierto grado de autonomía, con el fin de alcanzar objetivos específicos. Una de las ramas de la IA, es el Aprendizaje Automático o *Machine Learning* (ML). El ML diseña modelos predictivos que construyen por sí mismos la relación entre las variables a estudiar mediante el análisis de un conjunto inicial de datos, la identificación de patrones y el establecimiento de criterios de clasificación. Una vez fijados los criterios, al introducir un nuevo conjunto de datos el componente IA es capaz de realizar una inferencia. Por su parte, el algoritmo establece un sistema matemático de solución de problemas.

<sup>2</sup> El sector de asistencia sanitaria, de los seguros, sector bancario y financiero, la mercadotecnia y la publicidad son solo algunos ejemplos de los ámbitos en los que se lleva a cabo con más regularidad la elaboración de perfiles, para contribuir al proceso de toma de decisiones.

contratación y despidos. Un sistema algorítmico emplea la Agencia Estatal de Empleo en Austria (AMS) para la clasificación de los trabajadores y conocidas plataformas, como Glovo o Deliveroo, también se valen de un algoritmo para fijar los niveles de pago de los repartidores, determinando cómo se combinan la oferta y la demanda, cuál es el pago y cuáles son las consecuencias si un trabajador de la plataforma recibe calificaciones positivas o negativas<sup>3</sup>.

Pues bien, estas decisiones automatizadas tienen el potencial de afectar de forma significativa a los derechos y libertades de las personas y su utilidad para las organizaciones, al ofrecer respuestas eficientes con ahorro de costes, no debe ocultar sus riesgos para los derechos de las personas afectadas. La protección de las personas físicas, en relación con el tratamiento de datos personales es un derecho fundamental, al igual que el derecho a la intimidad y el derecho a la igualdad y la prohibición de discriminación, que deben seguir siendo garantizados frente a decisiones automatizadas con virtualidad para lesionarlos.

Sin embargo, algoritmo, sesgo y discriminación son términos que se conectan frecuentemente<sup>4</sup>. En efecto, un aspecto crítico de los sistemas de IA es el de la posible existencia de sesgos o “*bias*”, que serían desviaciones inadecuadas en el proceso de inferencia. Los sesgos son particularmente graves cuando, por ejemplo, derivan en discriminaciones de un grupo en favor de otro. Aunque el sesgo de la decisión no es un problema particular de los sistemas de IA, sino que es general de cualquier proceso de toma de decisión, ya sea humano o automático, sí es preocupante que la elaboración de perfiles pueda perpetuar los estereotipos existentes y ampliar la segregación social.

Como ya se ha señalado, decisiones de contratación laboral o de atribución de productividad con efectos salariales se desarrollan a través de algoritmos y lo preocupante es el sesgo discriminatorio de las mismas que, en algunos casos, ha sido noticia en conocidas empresas y objeto de estudios académicos, en muchos otros<sup>5</sup>. Como informaron los medios de prensa, Amazon tuvo que desmantelar en

<sup>3</sup> Vid STS de 25/09/2020, rec. 4746/2019.

<sup>4</sup> Vid., Zuiderveen Borgesius, F., (*Discrimination, artificial intelligence, and algorithmic decision-making*, Consejo de Europa, 2018, pp.11-14) sobre discriminaciones algorítmicas, en materia de política criminal preventiva, de selección de alumnos y de selección de trabajadores, publicidad, precios, sistemas de búsqueda a través de imágenes e incluso en el ámbito de las herramientas de traducción. Vid., Barocas S., Selbst, A., (Big Data’s Disparate Impact, *California Law Review*, 104/2016, p. 671) sobre casos de impacto adverso asociados a decisiones algorítmicas. Un trabajo clásico, con abundantes supuestos de hecho, es el de O’Neil, C., *Weapons of math destruction: How big data increases inequality and threatens democracy*, Crown, 2016.

<sup>5</sup> Si un algoritmo de contratación utiliza, en su búsqueda de los empleados más estables, el criterio predictivo de la distancia entre el lugar de trabajo y su domicilio, producirá un impacto racial desproporcionado, dado que en muchas ciudades la discriminación ha modelado patrones

2018 su sistema de contratación algorítmica, cuando tras un año de utilización se comprobaría que desfavorecía a las mujeres, pues el sistema prefería a los candidatos hombres. El sistema algorítmico de la Agencia Estatal de Empleo en Austria (AMS) presenta el riesgo de asignar las mujeres desempleadas a un grupo inferior, incluso si su experiencia y calificaciones coinciden con las de un hombre<sup>6</sup>. También, en relación con la edad como factor de discriminación, su relevancia en las plataformas de búsqueda de empleo está cada vez más documentada, particularmente, a los efectos de eliminar o disuadir a los solicitantes de empleo mayores de determinada franja de edad<sup>7</sup>. Igualmente, en diversos estudios se ha analizado detenidamente la discriminación salarial de las trabajadoras de plataformas; plataformas que usan algoritmos para fijar los niveles de pago<sup>8</sup>.

Ahora bien, partiendo de recordar que en el ámbito de la relación laboral el empleador está vinculado por la prohibición de discriminación, la cuestión que surge inmediatamente sería la de la efectividad de esta prohibición para evitar sistemas algorítmicos sesgados o con impacto adverso sobre determinados grupos de personas trabajadoras. A estos efectos, habría que plantear si puede hablarse de la existencia de una “discriminación algorítmica” distinta de la decisión humana discriminatoria, es decir, de una nueva categoría jurídica, cuyos rasgos diferenciales exijan la adaptación del marco normativo antidiscriminatorio.

## 2. DISCRIMINACIÓN ALGORÍTMICA: NUEVOS RASGOS Y NUEVOS RIESGOS

La primera dificultad para hacer frente a los efectos discriminatorios de las decisiones automatizadas radica precisamente en los rasgos que las caracterizan y que determinan nuevos riesgos y desafíos para la tutela antidiscriminatoria. La investigación jurídica debe partir de conocer los caracteres más destacados o sobresalientes, aquellos que las tecnologías digitales aportan al viejo problema de la discriminación en las relaciones laborales. A mi juicio, tres características

residenciales. Vid. Kim, P. T. Data-driven Discrimination at Work, *William. & Mary Law Review*, 58/2017, pp. 863 y 864.

<sup>6</sup> Vid., Allhutter D., Cech F., Fischer F., Grill G., Mager A., Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective, *Frontiers Big Data*, 21February 2020, <https://doi.org/10.3389/fdata.2020.00005>, p. 7. De otra parte, hemos de recordar en EEUU, el caso de *HireVue* que ofrece entrevistas virtuales con solicitantes de empleo. Esta plataforma utiliza una herramienta novedosa para la contratación que identifica la expresión facial, indicaciones de voz, elección de palabras. El problema estriba es que el sistema de análisis facial utilizado tiene dificultades para leer las caras de mujeres de piel oscura.

<sup>7</sup> Vid., Ajunwa, I., Age Discrimination by Platforms, *Berkeley Journal of Employment and Labor Law*, 40/2019, pp. 27 y 28.

<sup>8</sup> Vid., Kullmann M., Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law, *International Journal of Comparative Labour Law and Industrial Relations*, 34/2018, p. 20.

definitorias de la discriminación algorítmica son invisibilidad, intensidad y complejidad técnica.

El primer rasgo que caracteriza a la llamada discriminación algorítmica es su invisibilidad. Es así un riesgo invisible pues se afirma que el sesgo o el efecto discriminatorio sería común a cualquier decisión, (ya sea humana o automatizada); y en todo caso es mayor el riesgo discriminatorio de la decisión humana<sup>9</sup>. La evidencia de los efectos discriminatorios de las decisiones automatizadas no puede ocultarnos, se insiste, el sesgo en la toma de decisiones humanas. De hecho, la automatización de los procesos de toma de decisiones se presenta como una forma de evitar decisiones humanas subjetivas y sesgadas, al reemplazarlas por una técnica neutral. Conectado a este rasgo se encuentra la incuestionable creencia en la objetividad de los datos, que determina una aceptación acrítica de estos sistemas algorítmicos; proclamados como oráculos, consecuentemente, son aceptados sin análisis ni evaluación. Esta invisibilidad también se proyecta en la ausencia de responsabilidad humana sobre los resultados sesgados o discriminatorios, dada la opacidad de la participación humana en las decisiones automatizadas.

Un segundo aspecto novedoso de la discriminación algorítmica es su extensión e intensidad. Dado el elevado volumen, velocidad y variedad de datos usados, en cualquier proceso automatizado de contratación o decisión empresarial, cualquier sesgo introducido en el sistema magnificará y multiplicará sus efectos.

En consecuencia, la aplicación de la IA a la toma de decisiones no deja de ser, paradójica. Por un lado, tal automatización podría ser un intento de prevenir la discriminación, pero por otro lado, existe evidencia de que los procesos algorítmicos de toma de decisiones pueden frustrar los objetivos de las leyes antidiscriminatorias y servir para reproducir las desigualdades a mayor escala. En efecto, los sesgos existen en los sistemas algorítmicos, al igual que en el mundo real y estos sistemas reproducen o repiten las discriminaciones del mundo real. De una parte, estos efectos discriminatorios pueden producirse al margen de cualquier intención de discriminar pues género o raza pueden estar estadísticamente asociados con determinado criterio o rasgo, como la productividad del trabajador, de forma que un criterio admitido serviría para autorizar uno vedado. Pero, de otra parte, se afirma que tampoco debe descartarse totalmente la idea de que estos sistemas automatizados sirvan para “blanquear” decisiones humanas

<sup>9</sup> Bornstein, S., (Antidiscriminatory Algorithms, *Alabama Law Review*, 70/2018, p. 570) cuestiona el rigor de los estudios que muestran que las decisiones automatizadas son menos sesgadas que las decisiones humanas.

discriminatorias, configurando sistemas algorítmicos que excluyan determinados colectivos o minorías. Algunos autores describen mecanismos que permiten configurar procesamiento de datos discriminatorios y que podrían así blanquear decisiones intencionalmente discriminatorias<sup>10</sup>.

El tercer rasgo de la discriminación algorítmica es su complejidad técnica. En las aplicaciones de componentes de IA intervienen múltiples elementos técnicos<sup>11</sup>, pero también humanos, que hay que tener en cuenta para determinar sus implicaciones sobre los derechos de las personas. Para determinar estas implicaciones sobre los derechos (y en concreto sobre la prohibición de discriminación) será preciso saber cómo funciona la IA. Los sistemas algorítmicos no sólo se basan en una correlación o unión de datos sino que la interpretación está en el centro del análisis de estos datos. Pueden surgir sesgos de los datos, pero también de la llamada caja negra, es decir de la definición del algoritmo, de la forma en que éste trabaja y de su facultad de predicción, derivados de los sesgos presentes en la sociedad. Atendiendo a este rasgo de la complejidad técnica, abordar las exigencias de la tutela antidiscriminatoria frente a la discriminación algorítmica conduce, en primer lugar, en la dirección de promover o imponer el desarrollo de de sistemas menos sesgados y más justos.

Para concluir este apartado creo necesario insistir en la idea de que debemos huir de debates estériles conectados con los destacados rasgos de la discriminación algorítmica. Así, comparar decisiones humanas y decisiones automatizadas desde la perspectiva de su naturaleza o efectos discriminatorios; es decir, considero desaconsejable el debate en torno a si las decisiones empresariales basadas en procesos automatizadas producen un mayor o menor impacto discriminatorio que las basadas en la voluntad humana. Tampoco creo acertado contraponer las soluciones técnicas y las jurídicas frente a la discriminación algorítmica, afirmando que se trata solamente de un problema técnico, que se soluciona con algoritmos justos y equitativos. Las soluciones técnicas, a mi juicio, necesitarán para su implantación y generalización del impulso y promoción de las normas jurídicas.

A mi juicio, la pregunta correcta es si el marco normativo actual es adecuado para hacer frente a las decisiones automatizadas discriminatorias atendiendo a los rasgos descritos. El debate necesario debe enfocar las exigencias de actualización del marco normativo para hacer frente a los nuevos riesgos, derivados de aquellos rasgos diferenciales. Es preciso coordinar el enfoque técnico y jurídico para

<sup>10</sup> Vid., Barocas, S., Selbst, A., *Big Data's Disparate Impact*, cit., p.675. Ampliamente, sobre el tema, vid., Kim, P. T. *Data-Driven Discrimination at Work*, cit., p. 884.

<sup>11</sup> Puede relacionar estos tratamientos con Big Data, Internet de las Cosas (IoT), 5G/sistemas móviles, Edge Computing y Computación en la Nube. Vid., AEPD, *Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial*, Una introducción, Febrero, 2020, p. 7.

hacer frente a los riesgos discriminatorios de las decisiones automatizadas. Desde nuestra perspectiva jurídica, lo relevante será arbitrar un sistema de garantías que permita el conocimiento de estos sistemas algorítmicos, a los efectos de impedir que se adopten decisiones discriminatorias. Debemos centrarnos, en fin, sobre la necesaria coordinación entre el enfoque técnico y jurídico; sobre qué deberá hacerse transparente para que las decisiones automatizadas puedan ser revisables, y así garantizar también frente a ellas la prohibición de discriminación en el mundo del trabajo<sup>12</sup>. Analizaremos, a continuación, la efectividad del marco normativo de la prohibición de discriminación en el empleo y la ocupación y, a continuación, las diversas propuestas formuladas desde EEUU y desde la Unión Europea para la protección de la igualdad frente a las aplicaciones de la IA.

### **3. LA APLICACIÓN DE LA NORMATIVA ANTIDISCRIMINATORIA A LAS DECISIONES EMPRESARIALES AUTOMATIZADAS DISCRIMINATORIAS**

La normativa europea y española en materia de igualdad en la ocupación y el empleo<sup>13</sup> protege a las personas frente a discriminaciones directas e indirectas basadas en criterios vedados de diferenciación como sexo, raza, religión o convicciones, discapacidad, edad u orientación sexual y que, entre nosotros el art. 14CE extiende, ampliamente, a cualquier otra condición o circunstancia personal o social.

Como se sabe, existirá discriminación directa cuando una persona sea, haya sido o pudiera ser tratada de manera menos favorable que otra en situación análoga por alguno de los motivos antes mencionados y existirá discriminación indirecta cuando una disposición, criterio o práctica aparentemente neutros pueda ocasionar una desventaja particular a personas de los colectivos definidos por los anteriores motivos (sexo, raza, religión o convicciones, discapacidad, edad u orientación sexual), salvo que dicha disposición, criterio o práctica pueda

<sup>12</sup> Una IA fiable debe ser en primer lugar transparente y explicable y un primer inconveniente será jurídico, toda vez que el proceso de construcción de un algoritmo es confidencial, al estar amparado bajo la normativa de propiedad industrial.

<sup>13</sup> En materia de empleo y ocupación: Directiva 2000/78/CE de 27/11/2000 relativa al establecimiento de un marco general para la igualdad de trato en el empleo y la ocupación; Directiva 2006/54/CE de 5/7/ 2006, relativa a la aplicación del principio de igualdad de oportunidades e igualdad de trato entre hombres y mujeres en asuntos de empleo y ocupación (refundición); Ley Orgánica 3/2007, de 22 de marzo, para la igualdad efectiva de mujeres y hombres; Real Decreto Legislativo 1/2013, de 29 de noviembre, por el que se aprueba el Texto Refundido de la Ley General de derechos de las personas con discapacidad y de su inclusión social; Real Decreto Legislativo 2/2015, de 23 de octubre, por el que se aprueba el texto refundido de la Ley del Estatuto de los Trabajadores: Arts. 17, 22.3, 24.2, 28 y 85.1 y 2; Real Decreto Legislativo 3/2015, de 23 de octubre, por el que se aprueba el texto refundido de la Ley de Empleo: Arts. 33.4 f) y 35.

justificarse objetivamente con una finalidad legítima y salvo que los medios para la consecución de esta finalidad sean adecuados y necesarios<sup>14</sup>. Frente a la relevancia del elemento intencional en la discriminación directa, es su efecto negativo o *adverse impact* en el concepto norteamericano, sobre un colectivo tutelado y la inexistencia de justificación, los dos elementos de prueba decisivos, bajo el test de la discriminación indirecta.

Pues bien, algunos caracteres de la discriminación algorítmica como su objetividad (y autoridad matemática) y el ahorro de costes y tiempo en el ámbito de cualquier decisión empresarial, conllevarán problemas de prueba, tanto bajo del test de la discriminación directa como indirecta. Sus destacados rasgos de invisibilidad y complejidad técnica introducen elementos no considerados en la conceptualización de ninguna de estas formas de discriminación prohibida.

Atendiendo a los conceptos de discriminación directa o intencional y de discriminación indirecta o por su efecto adverso, este segundo concepto parecería, en principio, el más próximo a la denunciada discriminación algorítmica. Las decisiones automatizadas sesgadas que discriminen a personas de colectivos tutelados, por ejemplo, por razón de su género, raza o edad, lo más probable, en efecto, será que, entre nosotros, se analicen bajo el test de la prohibición de discriminación indirecta<sup>15</sup>.

Sin embargo, algunos análisis ya han afirmado que, aunque la prohibición de discriminación indirecta podría abarcar muchos de los efectos discriminatorios de la IA, la aplicación de esta prohibición será difícil<sup>16</sup>. En este sentido, hemos de destacar que la doctrina norteamericana se ha mostrado cauta a la hora de aplicar la normativa antidiscriminatoria en sus actuales términos. Cuando en la decisión empresarial ha jugado un papel los algoritmos de aprendizaje automático resulta particularmente difícil aplicar el canon de interpretación o el test del impacto adverso. La aplicación mecánica de la doctrina del *disparate impact* falla, en

<sup>14</sup> El art. 2 b ii) de la Directiva 2000/78/CE, respecto de las personas con una discapacidad determinada, establece que el empresario o cualquier persona u organización a la que se aplique esta Directiva, esté obligado, en virtud de la legislación nacional, a adoptar medidas adecuadas de conformidad con los principios contemplados en el art. 5, para eliminar las desventajas que supone esa disposición, ese criterio o esa práctica.

<sup>15</sup> En la doctrina norteamericana diversos autores han entendido que la discriminación algorítmica podría analizarse tanto desde la teoría del *disparate treatment* como del *disparate impact*. Sobre el tema, vid., Ajunwa I., (The paradox of automation as anti-bias intervention, *Cardozo Law Review*, 41/2020, p. 8) que afirma su “agnosticismo” sobre esta opción formulando un tercera vía.

<sup>16</sup> Vid., Zuiderveen Borgesius, F., *Discrimination, artificial intelligence, and algorithmic decision-making*, cit. p.19. Este autor también ve como debilidad de la normativa antidiscriminatoria para abordar la discriminación algorítmica que esta última se extiende a personas que no están en los colectivos amparados tradicionalmente por la tutela antidiscriminatoria.

efecto, para abordar los riesgos particulares de estos sistemas algorítmicos pues esta doctrina evalúa el uso de test o criterios, que miden habilidades de los trabajadores y se centra en la validez de tales medidas y su relevancia para un trabajo concreto. Por el contrario, los sistemas algorítmicos extraen los datos disponibles, buscando correlaciones estadísticas que conectan variables aparentemente no relacionadas, como los patrones de comportamiento de las redes sociales, con el desempeño laboral<sup>17</sup>. Los modelos algorítmicos no presentan los caracteres que se identifican para una aplicación correcta de la doctrina del *adverse impact*<sup>18</sup>, ya que, en lugar de proporcionar criterios de selección o de evaluación específicos que están justificados por su relevancia para el trabajo, los modelos de datos involucran procesos de decisión opacos, que descansan en correlaciones inexplicables, y carecen de justificaciones empresariales claramente articuladas.

En definitiva, habría que cuestionar esa fácil equiparación de la discriminación algorítmica a un eventual supuesto de discriminación indirecta. En efecto, debemos recordar que el concepto de discriminación indirecta significó la ampliación de la tutela antidiscriminatoria desde la discriminación intencional a la discriminación por efecto no justificada en la necesidad de la empresa. El concepto de discriminación indirecta está dirigido a evitar el uso de criterios neutros, pero con efectos discriminatorios, siempre que no estén justificados por su fin legítimo y no superen un escrutinio de adecuación y proporcionalidad. Sin embargo, cuando nos enfrentamos a una decisión automatizada, el problema supera o desborda el supuesto de hecho que llevaría a la ampliación de la prohibición de discriminación desde las discriminaciones directas a las indirectas. Nos enfrentamos a algo mucho más complejo, a un componente aplicativo de inteligencia artificial para adoptar decisiones parciales o finales automatizadas. Por ello, creo que el tema de la discriminación algorítmica debe abordarse con una nueva ampliación de la tutela antidiscriminatoria; una actualización del marco normativo sobre la igualdad materia de empleo y ocupación.

Como ya se ha afirmado, la aplicación de la IA a la toma de decisiones ha implicado un desafío legal sin precedentes, toda vez que la potencialidad de los sistemas algorítmicos para generar desigualdad y las dificultades para detectar y probar la discriminación de las decisiones empresariales basadas en algoritmos afectan a la efectividad de las leyes antidiscriminatorias<sup>19</sup>. Por ello, es preciso

<sup>17</sup> Vid., Kim, P.T., *Data-Driven Discrimination at Work*, cit., p. 907. Como esta autora señala, incluso se ha venido destacando, en términos más amplios, que la doctrina del *adverse impact* presenta problemas para ser aplicada fuera del contexto en que la misma se origina y desarrolla, el de criterios o prácticas fácilmente identificables.

<sup>18</sup> Vid., Selmi, M., *Was the Disparate Impact Theory a Mistake?*, *UCLA Law Review*, 53/2006, p. 705.

<sup>19</sup> Vid., Cohen, J. E., *Law for the Platform Economy*, *U.C. Davis Law Review*, 51/2017, p. 190

un cambio de estas “herramientas” legales, una extensión de las mismas a estas nuevas realidades tecnológicas. A este objetivo dedicaremos los siguientes apartados.

#### 4. PROPUESTA DE ACTUALIZACIÓN NORMATIVA EN ESTADOS UNIDOS

En la doctrina norteamericana se han formulado diversas respuestas legales frente a la discriminación algorítmica y a los análisis jurídicos que aplican el marco legal vigente<sup>20</sup>, se han contrapuesto los que abogan por actualizaciones y ampliaciones de los supuestos de discriminación prohibidas<sup>21</sup>. En este contexto, se formularía la llamada “teoría del sesgo de clasificación”. Según su autora, una lectura atenta de la normativa (Título VII *Civil Rights Act.*) sugiere que la misma prohíbe directamente las clasificaciones sesgadas. Esta clasificación sesgada se produce cuando los empresarios utilizan sistemas algorítmicos para seleccionar o clasificar (con diversas puntuaciones a trabajadores) que crean desventajas por razón de raza, sexo u otro criterio o característica protegida<sup>22</sup>. Esta teoría, reconoce la antes definida clasificación sesgada como un tipo separado de daño prohibido por la normativa antidiscriminatoria en EEUU<sup>23</sup>. En términos muy similares, se propone la denominada doctrina de la discriminación *per se* de los sistemas algorítmicos con efecto adverso sobre las categorías protegidas, y en la que trasladaría al empresario la carga de probar la necesidad de tales prácticas<sup>24</sup>. Esta doctrina parte de constatar, de un lado, que la línea de la responsabilidad (también empresarial) por los problemas creados por los algoritmos de aprendizaje automático es

<sup>20</sup> Vid., Baroca, S., Selbst, A., Big Data’s Disparate Impact, cit., pp. 671 y 674.

<sup>21</sup> Y al margen de estas existen también propuestas de regulación diversas, como sería la de desarrollar un control administrativo *ex ante*, es decir un régimen de regulación de los algoritmos como el que se usa actualmente para una aprobación previa a la comercialización de medicamentos: Vid Tutt, A., An FDA for Algorithms, *Administrative Law Review*, 68/2017, pp. 20-25.

<sup>22</sup> Vid., Kim, P.T., Data-driven Discrimination at Work, cit., p. 866. Se trata de una propuesta deducida del referido texto normativo y que está dirigida a establecer la responsabilidad empresarial y no del creador del sistema algorítmico pues aunque el citado Título VII de la *Civil Rights Act* se aplica a las agencias de colocación en difícil que pueda alcanzar a las compañías que proporcionen o vendan un sistema algorítmico de contratación.

<sup>23</sup> Teoría que aplicaría a la normativa de la UE, Kullmann M., Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law, cit., p. 18.

<sup>24</sup> Vid., Ajunwa, I. The paradox of automation as anti-bias intervention, cit., p. 45. Dadas las dificultades de prueba, esta nueva teoría de la responsabilidad empresarial modifica las reglas sobre la carga de la prueba, pues el Título VII de la *Civil Rights Act* requiere para establecer la responsabilidad o intención o, en ausencia de intención, un clara prueba del efecto adverso, sin que esté justificado en una necesidad empresarial. Se cita como antecedente de esta teoría que la *National Labor Relations Act* caracteriza algunas acciones empresariales como violaciones normativas *per se*, por ejemplo, un cambio adoptado de forma unilateral por el empresario sobre materias sujetas al deber de negociar.

borrosa y, de otro, que existen dificultades de prueba, para que pueda imputarse responsabilidad a la empresa, en caso de falta de intención de discriminar o de falta de certidumbre sobre los resultados sesgados de un algoritmo.

Y en la misma línea, otras propuestas fundamentan la responsabilidad del empresario, en caso de discriminación algorítmica, en su conducta imprudente. El empresario incurriría en responsabilidad derivada de imprudencia o negligencia, en caso de adoptar decisiones fundadas en sistemas algorítmicos con resultados sesgados, dada la posibilidad de prevenir y rectificar sesgos y estereotipos en las decisiones relativas la gestión de personal en los lugares de trabajo. Atendiendo al desarrollo de herramientas tecnológicas, los empresarios, que conocen el riesgo de los sesgos implícitos, (tienen evidencias sobre de las tendencias que pueden afectar sus decisiones) pueden utilizar algoritmos predictivos para reducir el sesgo de las decisiones, por lo que ante un fallo en materia de prevención, deben ser responsables por discriminación (intencional) derivada de su negligencia o imprudencia.

En definitiva, se afirma que de la prohibición de discriminar en el empleo se deriva un mandato legal de auditoría (interna o externa) de los sistemas de automatizados de decisión, pudiendo constituir la negativa del empresario a auditar y corregir un sistema que determina un sesgo ilegal una *prima facie* de un comportamiento discriminatorio<sup>25</sup>.

## 5. PROPUESTAS DESDE LA UNIÓN EUROPEA

La Agencia Europea de los Derechos Fundamentales (FRA) ha formulado también diversas propuestas para la protección de los derechos fundamentales en el desarrollo y uso de los algoritmos<sup>26</sup>. Estas propuestas debemos analizarlas en el marco de la Estrategia Europea en IA y en conexión con las soluciones ya avanzadas desde el Reglamento General de Protección de Datos (RGPD)<sup>27</sup>, que constituye, hasta la fecha, la regulación normativa más importante sobre el binomio derechos fundamentales e inteligencia artificial. Si la Estrategia Europea quiere situar a la persona en el centro de la IA, es decir persigue una IA centrada en el ser humano, el RGPD establece la prohibición general de decisiones únicamente

<sup>25</sup> Vid., Ford, R.T., Bias in the Air: Rethinking Employment Discrimination Law, *Stanford Law Review*, 66/2014, p. 1381. Bornstein, S., Reckless Discrimination, *California Law Review*, 105/2017 p. 1056.

<sup>26</sup> Vid., European Union Agency for Fundamental Rights (FRA) #BigData: Discrimination in data-supported decision making, 2018; European Union Agency for Fundamental Rights (FRA) Data quality and artificial intelligence, mitigating bias and error to protect fundamental rights, 2019.

<sup>27</sup> Reglamento 619/2016, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (RGPD).

automatizadas con efectos jurídicos sobre los interesados (art. 22) y les asegura una intervención humana y garantías como el derecho de explicación y la evaluación de impacto (arts. 13,14 y 15).

### **5.1. Estrategia Europea en IA: Marco ético respetuoso de derechos y valores**

La Estrategia Europea en IA de 25/4/2018<sup>28</sup>, el Plan Coordinado de 7/12/2018, las Directrices éticas sobre IA fiable de abril de 2019, la Resolución Parlamento Europeo de 12/2/2019 y el Libro Blanco sobre IA de 19/02/2020 son un buen ejemplo de que la IA se abordará a nivel europeo desde la perspectiva del respeto a los valores y derechos fundamentales de las personas. En este enfoque de una IA centrada en el ser humano, debe garantizarse un marco que promueva la innovación y además respete principios éticos, como la transparencia y la obligación de rendir cuentas. Ganar la confianza necesaria para que las sociedades acepten y utilicen la IA, exige que la tecnología sea predecible, responsable, verificable, respete los derechos fundamentales y observe reglas éticas.

Para hacer realidad una IA fiable, centrada en la persona y en el respeto de los derechos fundamentales, será necesario establecer exigencias de transparencia, explicabilidad y rendición de cuentas o “responsabilidad algorítmica”. Gracias a la transparencia y explicabilidad deberán identificarse y podrán eliminarse los sesgos discriminatorios, en la fase de recopilación de la información o en los propios métodos de desarrollo del sistema de IA (por ejemplo, la programación del algoritmo). Pero será necesario complementar los requisitos anteriores con la rendición de cuentas, que exige establecer mecanismos que garanticen la responsabilidad de los sistemas de IA y de sus resultados. A estos efectos, sería necesaria la fijación de auditorías independientes, que evalúen los algoritmos (los datos y los procesos de diseño).

### **5.2. El RGPD: garantía de las personas frente a las decisiones únicamente automatizadas**

La Estrategia Europea sobre IA se encuentra directamente inspirada por la regulación del RGPD relativa a las decisiones automatizadas. El Reglamento asegura una intervención humana en las decisiones basadas en procesos automatizadas (incluida la elaboración de perfiles) que determinen consecuencias o efectos jurídicos sobre las personas. Así, de partida, establece como regla una prohibición general de las decisiones “únicamente” automatizadas y, aunque prevé supuestos excepcionales en los que estas serían admisibles, los rodea de

<sup>28</sup> Comunicación de la Comisión Europea de 25/4/2018, sobre la Inteligencia artificial para Europa COM (2018) 237 final; Comunicación de la Comisión Europea de 7/12/2018 sobre un Plan coordinado sobre la inteligencia artificial COM (2018) 795 final.

garantías y derechos de la persona interesada, que están dirigidos a conocer la lógica del proceso decisional y sus efectos. En concreto, establece un “derecho de explicación” y la evaluación de impacto de la IA sobre la protección de datos, que son dos instrumentos relevantes para garantizar los derechos fundamentales.

El art. 22.1 RGPD garantiza el derecho a no ser objeto de una decisión basada “únicamente” en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en el interesado o le afecte significativamente de modo similar. En su interpretación de este art. 22.1 RGPD, el GT29<sup>29</sup> ha señalado que el precepto establece una prohibición general de las decisiones basadas únicamente en el tratamiento automatizado, lo que implicaría que las personas están protegidas automáticamente, frente a las mismas<sup>30</sup>. Estas decisiones exclusivamente automatizadas serían las que se adoptan sin una intervención humana “responsable”, que analice la información procesada y cuente con poder o capacidad de modificar la decisión automatizada<sup>31</sup>.

La protección del citado art. 22 RGPD se extiende a las decisiones únicamente automatizadas, tanto parciales o previas (elaboración de perfiles<sup>32</sup>) como finales o definitivas, pero siempre que produzcan efectos jurídicos o afecten significativamente de modo similar a la persona interesada; por tanto, esa afectación jurídica o significativa a los interesados de las decisiones automatizadas es otra pieza clave del precepto. De esta forma, la decisión debe tener el potencial de afectar significativamente a las circunstancias, al comportamiento o a las elecciones de las personas afectadas; tener un impacto prolongado o permanente en el interesado; o, en los casos más extremos, provocar la exclusión o discriminación de personas. Las

<sup>29</sup> Grupo de Trabajo de protección de datos, del art. 29 de la anterior Directiva 95/46/CE, derogada por el RGPD, actualmente sus funciones las desarrolla el Comité Europeo de Protección de Datos.

<sup>30</sup> Vid., Grupo de Trabajo sobre protección de datos del art. 29 (GT29), Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679, de 3/10/2017, revisadas el 8/2/2018, p. 21. Una interpretación que se apoya también en el Considerando 71 RGPD,.... “Sin embargo, se deben permitir...” y es seguida por la doctrina: Zuiderveen Borgesius, F., *Discrimination, artificial intelligence, and algorithmic decision-making*, cit., p.22; Todolí Signes, A., 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*, 84/2018, p.77.

<sup>31</sup> Vid., Grupo de Trabajo sobre protección de datos del art. 29 (GT29), Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679, cit., p.23.

<sup>32</sup> De conformidad con el art. 4.4 RGPD, la elaboración de perfiles es 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.

decisiones automatizadas, incluida la elaboración de perfiles, relativas a la contratación o despido de trabajadores o a su promoción económica o profesional estarían claramente dentro de las decisiones, a las que se refiere el art. 22 RGPD.

Sin embargo, se trata de una prohibición que no es absoluta, pues, excepcionalmente, se admiten tales decisiones, entre otros supuestos, si son necesarias para la celebración o ejecución de un contrato (art. 22.2). Así estarían autorizadas las decisiones “únicamente” automatizadas (incluida la elaboración de perfiles) con efectos jurídicos o similares en el ámbito laboral, que sean necesarias para la celebración y la ejecución del contrato, por ejemplo, cuando el volumen de datos y de personal afectado haga necesario el uso de aplicaciones de IA. En otro caso, estas decisiones automatizadas solo podrán actuar como colaboradores de los responsables empresariales, que analizarán la información procesada y podrán, en ciertos casos, modificar la decisión automatizada.

Finalmente, el RGPD establece importantes garantías frente a las decisiones automatizadas. Cuando se trata de decisiones únicamente automatizadas con efectos relevantes sobre las personas interesadas, el RGPD establece una regulación específica, tanto en relación con los derechos de información de los interesados (arts. 13 y 14), como en relación con el derecho de acceso de estos (art. 15). El responsable debe informar de la existencia de estas decisiones automatizadas, incluida la elaboración de perfiles y, al menos en tales casos, ofrecer información significativa sobre la lógica aplicada, así como sobre la importancia y las consecuencias previstas de dicho tratamiento para el interesado (arts. 13.2.f), 14.2.g) y 15.1.h)<sup>33</sup>. La importancia de este derecho aconseja el tratamiento separado, en el próximo apartado.

### **5.3. El derecho de explicación sobre las decisiones basadas únicamente en procesos automatizados**

Los artículos 13.2.f) y 14.2.g) del RGPD configuran el llamado derecho de explicación sobre la lógica aplicada por las decisiones exclusivamente automatizadas (así como sobre la importancia y las consecuencias previstas de dicho tratamiento para el interesado), que deberá realizarse a través de una información específica y de fácil acceso. El RGPD exige que el responsable del tratamiento ofrezca información significativa sobre la lógica aplicada; no necesariamente una compleja explicación de los algoritmos utilizados o la revelación de todo el algoritmo, pero sí una información suficientemente exhaustiva para que el interesado entienda los motivos de la decisión<sup>34</sup>.

<sup>33</sup> Vid. art. 22.3 RGPD.

<sup>34</sup> En sus recomendaciones, el citado GT29 indicaba que, en lugar de ofrecer una compleja explicación matemática sobre cómo funcionan los algoritmos o el aprendizaje automático, el res-

Pues bien, de una parte, no puede desconocerse que la efectividad de este derecho de explicación ha sido puesta en duda por quienes afirman que los algoritmos contemporáneos se basan en modelos, que exhiben una lógica implícita, más que explícita, por lo que la lógica subyacente a su toma de decisiones representaría una “caja negra” ininterpretable<sup>35</sup>. Sin embargo, de otra parte, otros autores destacan el significativo cambio introducido por el RGPD en relación con el poder de control y sanción del incumplimiento de esta obligación por parte de la autoridad de control (arts. 53 y 83). Por ello, se concluye que el derecho de explicación se comprende mejor a la luz de los citados preceptos, que reconocen a la autoridad de control poderes de investigación y sanción necesarios para garantizar la efectividad del derecho a la explicación de las personas afectadas por decisiones automatizadas<sup>36</sup>.

Pues bien, estas garantías y derechos del RGPD en favor de la persona interesada resultan aplicables en el caso de una empresa utilice sistemas automatizados de decisión, incluida la elaboración de perfiles, por ejemplo, para la contratación de sus trabajadores. La empresa estaría obligada así, en primer lugar, a informar sobre la existencia de una decisión únicamente automatizada y el supuesto excepcional en el que se fundamenta su admisión, dada la prohibición general del art. 22.1 RGPD<sup>37</sup>. Además, resultaría aplicable el denominado derecho de explicación sobre la lógica aplicada por el sistema algorítmico, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado. Si estamos ante decisiones únicamente automatizadas (aunque sean parciales)

ponsable del tratamiento debe considerar la utilización de formas claras y exhaustivas de ofrecer información al interesado, por ejemplo: las categorías de datos que se han utilizado o se utilizarán en la elaboración de perfiles o el proceso de toma de decisiones; por qué estas categorías se consideran pertinentes; cómo se elaboran los perfiles utilizados en el proceso de decisiones automatizadas, incluidas las estadísticas utilizadas en el análisis; por qué este perfil es pertinente para el citado proceso; y cómo se utiliza para una decisión relativa al interesado. Vid., Grupo de Trabajo sobre protección de datos del art. 29, Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679, cit., p. 28.

<sup>35</sup> Vid., Goodman, B., Flaxman, S., EU Regulations on Algorithmic Decision Making and “a Right to an Explanation,” *ICML Workshop on Human Interpretability in ML*, 2016; Wachter S., et al., Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation, *International Data Privacy Law* 7/2017, pp. 76-99.

<sup>36</sup> Téngase en cuenta que las multas previstas por el incumplimiento de esta obligación de informar sobre la lógica de la decisión automatizada puede alcanzar hasta los 20 millones de euros o el 4% del volumen de negocio anual del ejercicio anterior, optándose por la de mayor cuantía, lo que determina que ninguna empresa sea inmune al cumplimiento. Vid., Casey, B., The Next Chapter in the GDPR’s “Right to Explanation” Debate and What It Means for Algorithms in Enterprise, *European Union Law Working Papers*, Stanford-Vienna Transatlantic Technology Law Forum, 2018, p. 42.

<sup>37</sup> Y en su caso, que el proceso de contratación no está basado únicamente en decisiones automatizadas, identificando el grado o nivel de participación de la decisión humana.

habrá de garantizarse al interesado, el derecho a obtener una intervención humana por parte del responsable, a expresar su punto de vista y a impugnar su decisión. Y finalmente, la empresa estaría bajo los poderes de investigación y sanción de las autoridades de control de protección de datos, entre nosotros de la Agencia Española de Protección de Datos. En esta correlación de derechos individuales y de poderes o competencias de la autoridad de control se encontraría la garantía de la efectividad del derecho de explicación de las personas afectadas por decisiones automatizadas.

Desde una perspectiva dogmática, es preciso destacar que, a través de la aplicación de las garantías establecidas por el RGPD, la libertad sustantiva y formal que caracteriza algunas decisiones de gestión empresarial, como es el caso de las decisiones en materia de contratación o la política retributiva empresarial (en cuanto decisiones no sometidas a causa ni a procedimiento) resultaría limitada, en el caso de la introducción de aplicaciones de IA en la toma de estas decisiones por parte de la empresa.

Por ejemplo, la utilización de sistemas algorítmicos como colaboradores de los procesos de contratación de trabajadores, reduciría la discrecional empresarial y podría, consecuentemente, reducir su sesgo o efectos discriminatorios. De esta forma, ciertamente, la introducción de aplicaciones de IA en las decisiones empresariales de contratación, interesa a la empresa por el ahorro de costes y de tiempo y por las amplias posibilidades de evaluación de múltiples candidatos para contratar a los trabajadores más adecuados. Pero, al mismo tiempo, no deja de ser paradójico que la incardinación de las mismas dentro del ámbito de aplicación del RGPD determine necesariamente la “procedimentalización” de esta facultad empresarial, al quedar sometida a exigencias de transparencia, explicabilidad y evaluación de impacto. Las personas candidatas, participantes en los procesos de contratación automatizados, tendrán garantizados, de conformidad con el citado RGPD, los derechos de explicación, intervención humana, evaluación, revisión e impugnación, analizados. La autoridad con competencia de control en protección de datos, desarrollará sus competencias de investigación y sanción, en caso de denuncias de incumplimiento de aquellas garantías. El intenso control de las autoridades competentes y las eventuales (y significativas) sanciones pecuniarias, establecidas por el Reglamento, podrán garantizar que en las empresas los procedimientos de contratación de las personas trabajadoras se rijan por los principios de transparencia y dación de cuentas.

#### **5.4. Propuestas de protección de los derechos fundamentales en el desarrollo y uso de algoritmos**

La Agencia Europea de los Derechos Fundamentales ha elaborado propuestas de protección de los derechos fundamentales en el desarrollo y uso de

algoritmos que deben ser atendidas. En primer lugar, se apela, a la exigencia de transparencia sobre el proceso de elaboración de los algoritmos, para desarrollar estas herramientas y permitir detectar y, por tanto, rectificar cualquier error en su aplicación. En segundo lugar, se aboga por la realización de una evaluación del impacto de los derechos fundamentales, para identificar posibles sesgos y abusos en la aplicación y en los resultados obtenidos por los algoritmos. Esta evaluación incluye, entre otras, una medición del “potencial de discriminación”, en relación con motivos diferentes, como género, edad, etnia, religión, orientación sexual y política<sup>38</sup>. Transparencia y evaluación deben verificarse tanto sobre la calidad de los datos, como sobre la forma en que funcionan los algoritmos (caja negra) y las facultades predictivas de los mismos

A estos efectos, es esencial la comprobación de la calidad de los datos<sup>39</sup>. Dada la cantidad de datos generados y utilizados, sigue siendo un desafío evaluar la calidad de todos los datos recopilados y utilizados para construir algoritmos. Es esencial recopilar metadatos (es decir, información sobre los datos) y hacer evaluaciones de calidad de la corrección y generalización de los datos<sup>40</sup>. Aunque estas evaluaciones sobre calidad de datos pueden ayudar a comprender y mitigar los efectos discriminatorios de estos sistemas, la evaluación de los algoritmos no puede detenerse aquí; debe extenderse al desarrollo de los algoritmos, a la forma en que trabajan (la llamada caja negra) y a los poderes de predicción de los mismos.

<sup>38</sup> Al mismo tiempo, las evaluaciones de impacto podrían evaluar el posible sesgo de discriminación derivado de usar “proxy” información, como por ejemplo puede ser el lugar de residencia, con respecto a motivos protegidos, raza, en el área de discriminación. Vid., European Union Agency for Fundamental Rights (FRA) #BigData: Discrimination in data-supported decision making, 2018.

<sup>39</sup> Como ya se destacó, pueden surgir sesgos en la forma de recopilar los datos y de escribir el algoritmo que pueden derivarse de sesgos presentes en la sociedad. Por tanto, la calidad de los datos, (datos incompletos y sesgados puede conducir a resultados discriminatorios) junto con el diseño de los algoritmos y los procesos de reevaluación constante, deberían evitar la aparición de sesgos. La transparencia sobre los datos utilizados en cada sistema ayudará a prevenir la violación de derechos, lo que especialmente importante en casos de *big data*, donde el volumen de datos se valora más que su calidad.

<sup>40</sup> En su propuesta para evaluar la calidad de los datos en las aplicaciones de *Machine Learning*, las preguntas que pueden ayudar a identificar si debido a la calidad de los datos existen potenciales problemas de impacto adverso sobre derechos fundamentales con el uso de un algoritmo serían: 1) ¿De dónde provienen los datos? ¿Quién es responsable para recopilación de datos, mantenimiento y diseminación? 2) ¿Qué información se incluye en los datos? ¿Es la información incluida en los datos apropiados para el propósito del algoritmo? 3) ¿Quién está cubierto en los datos? ¿Quién está subrepresentado en los datos? 4) ¿Falta información dentro del conjunto de datos? ¿O hay algunas unidades solo parcialmente cubiertas? 5) ¿Cuál es el marco temporal y la cobertura geográfica de la recopilación de datos utilizada para construir la aplicación? Debe recordarse que, de momento, no existen estándares acordados sobre la evaluación de calidad de datos para las aplicaciones de *Machine Learning*.

La evaluación de los algoritmos, desde una perspectiva de derechos fundamentales, es una tarea compleja que exige una evaluación global, que incluya los efectos del uso de los datos sobre las libertades y derechos fundamentales, como el principio de no discriminación<sup>41</sup>. Es preciso, pues, una evaluación holística sobre el impacto en derechos fundamentales, en la que la evaluación de los procesamientos de datos debe ser solo una parte. En definitiva, se propone que la protección de los derechos fundamentales frente a la IA se articule en base a exigencias de transparencia y explicabilidad, y una evaluación amplia del impacto sobre los derechos fundamentales. Estas exigencias serían aplicables, por tanto, frente a las discriminaciones algorítmicas.

## 6. A MODO DE RECAPITULACIÓN Y PROPUESTAS PARA ESPAÑA

En las propuestas analizadas, se aprecia un elemento común; la consideración de la discriminación algorítmica como un riesgo frente al que ha de establecerse un marco normativo de tutela preventiva. Más concretamente, centrando su aplicación al ámbito de la empresa y a su actuación en el terreno de la decisión empresarial, esta tutela preventiva frente al riesgo de la discriminación algorítmica impondría al empresario una evaluación de impacto o una auditoría interna para medir y corregir, en su caso, el impacto discriminatorio de las decisiones automatizadas. Las exigencias de transparencia, explicabilidad y evaluación o auditoría algorítmica, sobre las que se insiste desde todas las propuestas indicadas, son ciertamente adecuadas si recordamos los rasgos de invisibilidad, intensidad y complejidad técnica que definían este nuevo riesgo. Sólo a través de la transparencia y explicabilidad podrá evaluarse si el impacto discriminatorio deriva de la (falta de) calidad de los datos, de la forma en que trabajan (la llamada caja negra) o/y de los poderes de predicción de los algoritmos. Sólo así será posible una evaluación holística que mitigue o rectifique el referido impacto.

Configuradas como principios éticos por la Estrategia Europea en IA, la transparencia, explicabilidad y la evaluación o auditoría algorítmica, son las exigencias del estándar de regulación establecido por el RGPD, que constituye un modelo para la normativa antidiscriminatoria. Tras el análisis de la regulación del RGPD puede constatar que la misma presenta carencias, como herramienta frente a la discriminación algorítmica y que, entre otras causas, se han derivado de su déficit de cumplimiento, de su limitado alcance<sup>42</sup>, o del carácter abierto y

<sup>41</sup> Vid., Mantelero, A., AI and Big Data: A blueprint for a human rights, social and ethical impact assessment, *Computer Law and Security Review*, 34/2018, p. 772

<sup>42</sup> Su limitado alcance deriva de que parte de los procesos algorítmicos están fuera del alcance de esta normativa de protección de datos, que solo se aplica cuando se procesan datos personales.

abstracto de sus normas<sup>43</sup>. En cualquier caso, aunque ofrece posibilidades para combatir la discriminación, tal vez sea aún demasiado pronto para evaluar el efecto del RGPD como una herramienta frente a la discriminación<sup>44</sup>. En consecuencia, resulta necesaria la adaptación de la normativa antidiscriminatoria sobre la base de los estándares previstos por el citado RGPD, que convergen con las propuestas de la doctrina de EEUU para el control de las decisiones algorítmicas sesgadas y articuladas, igualmente, en base a instrumentos o herramientas de transparencia, información y auditoría algorítmica.

Sobre la base de todo lo anterior, la propuesta para la regulación española nos llevaría a introducir en nuestra normativa laboral la obligación legal del empresario de evaluar el impacto discriminatorio de las decisiones automatizadas incluida la elaboración de perfiles, con efectos sobre las personas trabajadoras, en el empleo y en las condiciones de trabajo, a través de una reforma del art. 17 LET. Además, la relevancia colectiva del impacto discriminatorio de las decisiones empresariales ha de ser atendida mediante la acción de los sujetos de representación y tutela del interés colectivo, que son los representantes sindicales. Deben ser los representantes legales de los trabajadores, por tanto, los titulares del derecho de información y explicación sobre los sistemas algorítmicos de contratación o de gestión de personal y a los que la norma debe, reformándose el art. 64LET, igualmente, garantizar competencias de participación y control sobre las auditorías algorítmicas. Finalmente, podríamos plantear el debate sobre la oportunidad legal de fomentar un cumplimiento de esta obligación que excluya la responsabilidad empresarial, lo que significa la promoción de algoritmos justos y equitativos. De hecho, resultaban esperanzadoras las palabras de Bryce Goodman, cuando afirmaba que “en el pasado, las empresas han dedicado inmensos recursos para mejorar el algoritmo de actuación y, en el futuro, esperamos ver inversiones similares en la promoción de algoritmos justos y responsables”<sup>45</sup>.

No se aplica a los modelos predictivos porque no se relacionan con personas identificables. Vid. Zuiderveen Borgesius, F., *Discrimination, artificial intelligence, and algorithmic decision-making*, cit., p. 24.

<sup>43</sup> Además la normativa de protección de datos tiene reglas estrictas sobre “categorías especiales” de datos (a veces llamados “datos confidenciales”), que las organizaciones no podrán pues utilizar, y ello conlleva un desafío para evaluar y mitigar la discriminación, toda vez que los métodos para abordar la discriminación en los sistemas de IA suponen implícitamente que las organizaciones tienen estos datos confidenciales. Por último, como ya se vió, incluso cuando las explicaciones de las decisiones de IA pueden ser legalmente requeridas por el RGPD (o el Convenio 108+) a menudo es difícil explicar la lógica detrás de una decisión, cuando un sistema de IA, que analiza grandes cantidades de datos, llega a esa decisión.

<sup>44</sup> Se necesita más investigación legal sobre cómo la ley de protección de datos podría ayudar a mitigar los riesgos de discriminación: Vid., Zuiderveen Borgesius, F., op. cit., p. 24.

<sup>45</sup> Citado por Casey, Bryan, *The Next Chapter in the GDPR’s “Right to Explanation” Debate and What It Means for Algorithms in Enterprise*, cit., p. 56.

Si el proceso de la IA es imparable, es preciso regularlo desde la protección de la igualdad y la promoción de la diversidad como valor, también dentro de nuestras empresas, y por tanto promover el uso de algoritmos justos y equitativos. De esta forma, la intervención normativa, además de conjurar el riesgo de que los sistemas algorítmicos perpetúen los sesgos y la discriminación en el empleo y la ocupación puede convertirlos en la oportunidad de avance de la igualdad y la diversidad en el mundo de la empresa<sup>46</sup>.

<sup>46</sup> El proyecto a nivel europeo FA\*IR se basa en la aplicación de un algoritmo para frenar la discriminación en los procesos de búsqueda de empleo; su objetivo es la creación de un motor de búsqueda justo que persiga los agravios producidos en los portales digitales de empleo, rectificando su funcionamiento en pro de la igualdad. Hay que tener en cuenta que los algoritmos pueden aprender patrones discriminatorios y generar efectos sistemáticamente negativos sobre colectivos concretos, ya que no están programados para corregir prejuicios humanos. Frente a esta realidad, FA\*IR es capaz de crear un algoritmo de búsqueda que, sin afectar a la validez del ranking, podrá reorganizar los resultados con un mecanismo de acción positiva que detecte y relegue los mecanismos discriminatorios, incapacitando el sesgo étnico, de género o de edad.