Recibido: 7.1.2022. Aceptado: 17.03.2022

"Sesgos de género en el uso de inteligencia artificial para la gestión de las relaciones laborales: análisis desde el derecho antidiscriminatorio"

"Gender biases in the use of artificial intelligence for the management of labour relations: analysis from the anti-discrimination law"

Pilar Rivas Vallejo

Catedrática de Derecho del Trabajo y de la Seguridad Social

Universidad de Barcelona

pilar.rivas.vallejo@ub.edu

ORCID iD 0000-0002-1766-7659

RESUMEN

El análisis jurídico de la discriminación derivada de decisiones automatizadas que puedan provocar un impacto discriminatorio requiere combinar dos campos jurídicos: el de la protección de datos y el derecho antidiscriminatorio. En el primero los derechos reconocidos son accesorios al núcleo principal de afectación: el derecho de intervención humana y, principalmente, la explicabilidad de los algoritmos, manifestación de la debida justificación objetiva y razonable que acompaña a las decisiones *prima facie* discriminatorias para eludir su calificación como tales. Pero el tratamiento jurídico de la discriminación algorítmica requiere, también, dar respuesta a problemas de calificación de los sesgos en los que incurre el aprendizaje automático como resultado de las infinitas inferencias de datos que perfilan a personas en el contexto del derecho antidiscriminatorio, donde potencian su impacto discriminatorio, como son la discriminación por asociación o la discriminación múltiple o interseccional.

PALABRAS CLAVE: sesgos algorítmicos, discriminación, interseccionalidad, explicabilidad, intervención humana significativa.

ABSTRACT

The legal analysis of discrimination derived from automated decisions that may have a discriminatory impact requires combining two legal fields: data protection law and anti-discrimination law. The recognized rights of first field are accessory to the nucleus of affectation, and they're the right of human intervention and, mainly, the explicability of the algorithms, which constitutes a manifestation of the due objective and reasonable justification in case of discriminatory prima facie decisions in order to elude their qualification as such. But the legal approach to algorithmic discrimination also requires responding to problems of qualifying the biases derived from machine learning as a result of the infinite data inferences that outline people in the context of anti-discrimination law, where they enhance their discriminatory impact, such as discrimination by association or multiple or intersectional discrimination.

KEYWORDS: algorithmic biases, discrimination, intersectionality, explicability, significant human intervention

SUMARIO

I. VIEJOS PREJUICIOS CON NUEVA APARIENCIA

Para O'Neil[1], los algoritmos son modelos matemáticos incontestables, secretos e injustos. si bien "es crucial entender que, bajo la apariencia de neutralidad de los algoritmos, hay decisiones morales que perpetúan y aumentan las desigualdades sociales" (por ello los denomina "armas de destrucción matemática"). Los viejos prejuicios sociales se han digitalizado en la sociedad del siglo XXI; ahora se hacen visibles por medios que pueden reforzar los prejuicios sociales clásicos y proyectar exponencialmente su impacto, especialmente en el mundo del trabajo[2].

---

[1] O'Neil, C.: *Armas de destrucción matemática*, ed. Capitán Swing, Madrid, 2017.

[2] Rosenblat, A.: Uberland *Cómo los algoritmos están reescribiendo las reglas de trabajo*. University of California Press, 2018; y Umoja Noble, S.: *Algorithms of oppression: how search engines reinforce racism.* NYU Press, 2018. Vid., asimismo, Angwin, J., Larson, J., Mattu, S. y Kirchner, L.: "Machine

En efecto, el escenario que plantea el uso generalizado de algoritmos como la mano invisible e *irresponsable* que determina los procesos de selección de personal o las condiciones de trabajo, así como la propia continuidad en la empresa, pone en riesgo el valor de la igualdad y el derecho a la no discriminación (singularmente de las mujeres y de ciertos colectivos tradicionalmente excluidos o relegados del mercado de trabajo). La falta de transparencia u opacidad propias de los mecanismos automatizados incrementa las habituales dificultades en la reclamación contra las decisiones empresariales, lo que ha generado una creciente preocupación por el tema en el ámbito jurídico[3], aunque no es nueva en el área de la minería de datos o del aprendizaje de máquina y la inteligencia artificial (donde hace casi tres lustros existe un debate sobre el impacto de los algoritmos basados en datos como origen de sesgos sistemático contra grupos de personas, causado en parte por patrones históricos y actuales de discriminación, demostrando que personas que ya parten de una situación de desventaja pueden resultar en aún mayor desventaja como consecuencia del uso de un algoritmo[4]).

Estos sesgos ocultos en ocasiones obedecen a un objetivo claramente discriminatorio (discriminación directa), pero, en la mayoría de los casos, simplemente son provocados por el simple desinterés hacia su impacto colateral. La hipotética asepsia del algoritmo se presenta como un mecanismo alternativo a los prejuicios o sesgos humanos. Pero, si ese hipotético modelo de objetividad se basa en datos históricos "reales" (captando conductas discriminatorias reales, cuya reiteración en el tiempo detectada por el algoritmo conduzca a que este la identifique como la decisión correcta), su propio diseño escapa del criterio de la objetivación pretendida, sustituido por el de la eficiencia y la productividad. Y, al mismo tiempo, cuando se usa para adoptar decisiones laborales, puede servir para eludir la eficacia del plan de igualdad de las empresas, cuyas medidas se han podido construir precisamente para superar esas situaciones previas que sirven de soporte al modelo automatizado de decisión, lo que implica que, desde la perspectiva de la igualdad entre mujeres y hombres, la progresiva sustitución de mecanismos tradicionales de decisión por estos automatizados asistidos por inteligencia artificial juega claramente en contra de medidas pactadas para superar las desigualdades relativas al acceso al empleo y a las condiciones de trabajo[5], a menos que el propio plan de igualdad prevea mecanismos de corrección del impacto de herramientas automatizadas, que deberían ser centrales en el diseño de dichos planes, donde hasta el momento no encuentran reflejo alguno, ni en el plano legal ni en el plano práctico. De igual modo, el análisis de impacto de riesgos que debe aplicarse a la

---

bias: There's software used across the country to predict future criminals. and it's biased against blacks", 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[3] En este tema han sido pioneras las aportaciones de Pasquale y de Barocas-Selbst: Pasquale, F.: "A rule of persons, not machines: the limits of legal automation", *The George Washington Law Review*, vol. 87, núm. 1, 2019; Pasquale, F.: *New laws of robotics: defending human expertise in the age of AI*. The Belknap Press. 2020; y Barocas, S. y Selbst, A. D.: "Big data's disparate impact". *California Law Review*, núm. 104, 2016, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899. Se recomiendan en particular los trabajos de R. Xenidis citados en otras notas, así como de Zuiderveen Borgesius, F.: *Discrimination, artificial intelligence, and algorithmic decision-making*. Council of Europe, 2018.

[4] Hajian, S., Bonchi, F. y Castillo, C.: "Algorithmic bias: from discrimination discovery to fairness-aware data mining", DOI: 10.1145/2939672.2945386, 22nd ACM SIGKDD International Conference, 2016, https://www.researchgate.net/publication/305997939_Algorithmic_Bias_From_Discrimination_Discovery_to_Fairness-aware_Data_Mining.

[5] Se remite su tratamiento al estudio exhaustivo de Rivas Vallejo, P.: *La aplicación de la inteligencia artificial al trabajo y su impacto discriminatorio.* Aranzadi, Cizur Menor, 2020.

introducción de tales sistemas de gestión de personal haya de incluir el impacto de género, y su examen por parte de la comisión de seguimiento del plan de igualdad. Sin duda estos elementos debieran formar parte de los requisitos legales previstos en el RD 901/2021, de 13 de octubre, por el que se regulan los planes de igualdad[6].

La aparente neutralidad, objetividad y asepsia de los mecanismos automatizados de decisión juega sin duda en contra de la tutela del derecho a la igualdad, por cuanto la técnica del aprendizaje profundo en la que consisten, a partir de los datos de alimentación por parte de algoritmos que autoaprenden de ellos e infieren conclusiones, utilizadas para asesorar decisiones, impide establecer una clara conexión entre tales datos de alimentación (macrodatos) y las conclusiones a las que llega el modelo matemático. La perversidad de este funcionamiento es, precisamente, la dificultad de conocer dónde se encuentra el sesgo de la decisión si esta es discriminatoria (v.g., en un proceso de selección, por qué ha decidido que un individuo tiene "más valor laboral" que otro), y detectar el error a fin de impugnar la decisión empresarial.

II. SESGOS ALGORÍTMICOS BAJO EL PRISMA DEL DERECHO

A. Lenguaje jurídico y lenguaje de computación: su necesaria equivalencia

1. Interrelaciones y equivalencias entre ambos lenguajes

Tradicionalmente, los científicos de la computación y la inteligencia artificial se referían a la existencia de sesgos en la inteligencia artificial (IA), que asociaban a diversas tipologías, y que conectaban a un fenómeno holístico donde se integran diversas concepciones, afirmando que, pese a todo, las herramientas basadas en datos siempre serán mucho más exactas que los juicios emitidos por profesionales[7]. La incorporación de los juristas a este debate genera una interesante sinergia entre ambos campos del conocimiento, que aporta el imprescindible análisis jurídico del impacto discriminatorio de los algoritmos. Desde esta perspectiva, se ha abordado incluso la traducción del lenguaje computacional al jurídico, cuyas interrelaciones se revelan asimismo como necesarias tanto para la automatización de la aplicación del derecho y de la justicia como para el tratamiento jurídico de los sesgos algorítmicos. Y es que, como subraya Hildebrandt[8], el lenguaje de computación podría "erosionar la gramática

---

[6] La modificación de algoritmos en orden a reducir o esquivar los sesgos es factible mediante mecanismos técnicos de corrección que eviten su aprendizaje libre. Investigadores de las áreas de minería de datos, aprendizaje de máquina e inteligencia artificial, han realizado ya algunas propuestas correctivas, centradas en cada caso en una fase de diseño o aplicación del algoritmo, entre ellos, Hajian, Bonchi y Castillo (cit. en nota anterior). Otros autores ya han trabajado en el binomio algoritmo-discriminación desde una perspectiva más sociológica o jurídica (cfr. Smith-Strother, L.: "The role of social advocacy in diversity & inclusion recruiting", Glassdoor Summit, 2016, https://youtu.be/IdsqQMV4V_0), y otra línea interesante en el ámbito europeo ha incidido en la perspectiva de género (cfr. Barzilay, A. y Ben-David, A.: "Platform inequality: gender in the gig-economy", *Seton Hall Law Review*, vol. 47, núm. 2, 2017, http://scholarship.shu.edu/shlr/vol47/iss2/2; o Xenidis, R. y Senden, L.: "EU non-discrimination law in the era of artificial intelligence: mapping the challenges of algorithmic discrimination". U. Bernitz et al (eds): *General principles of EU law and the eu digital order*. Kluwer Law Int., 2020, pp. 151-182).

[7] Grove, W. M., et al.: "Clinical versus mechanical prediction: a meta-analysis". *Psychological assessment*, vol. 12, núm. 1, 2000.

[8] Hildebrandt, M.: "Algorithmic regulation and the rule of law". *Philosophical Transactions of the Royal Society A*, vol. 376, núm. 2128, 2018, DOI: http://dx.doi.org/10.1098/rsta.2017.0355.

y el alfabeto del derecho positivo moderno", lo que requerirá una nueva hermenéutica que exige la adecuada comprensión del vocabulario y la gramática del aprendizaje automático. En el análisis del impacto de los sesgos, la literatura especializada apunta también a la divergencia de lenguaje (Chouldechova[9] sostiene que la noción de discriminación indirecta no es un concepto estadístico en el sentido matemático, sino ético-jurídico, puesto que un instrumento técnico libre de sesgos predictivos puede provocar una discriminación indirecta en función del contexto al que se aplique), y equipara justicia con inexistencia de sesgo.

Del mismo modo, el concepto de "fairness" o justicia que se emplea en el campo de la computación para el análisis de sesgos, pudiendo tener una más acotada correspondencia con el de equidad, dista de ajustarse al lenguaje jurídico, toda vez que, en el ámbito del derecho antidiscriminatorio, son otros conceptos los realmente determinantes, como es el caso del equilibrio entre tal impacto y la justificabilidad, proporcionalidad y racionalidad de la decisión, no equiparables propiamente al principio de justicia material en la aplicación del derecho (que debe ponderarse en la aplicación de las normas en atención a las circunstancias del caso, según el art. 3.2 del C.c.[10]). La Resolución del Parlamento Europeo, de 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley (2016/2225(INI)), alude en su apartado 22 (referido al impacto discriminatorio de los sistemas de algoritmos y conjuntos de datos) a la equidad como criterio que debe guiar el examen de las predicciones basadas en el análisis de datos. Como instrumento de corrección de la ley en su ajuste a las circunstancias concretas del caso específico, introduciendo en su aplicación los criterios informadores de los principios generales del derecho (impregnados de un aura ética[11]), la equidad implica igualdad e individualización de trato, por lo que obliga a tratar igual a los casos iguales y desigual a los desiguales, combinando la epiqueya aristotélica y la aequitas romana como "atributo ético-jurídico del derecho"[12].

En todo caso, es insoslayable la necesidad de aproximar ambos lenguajes[13], para que la ciencia de la computación incorpore los conceptos jurídicos del contexto para el que se diseña o debe actuar, o, al menos, un lenguaje jurídico común que permita su adaptación a diferentes ámbitos locales o regionales. Pese a ello, existen cuestiones derivadas de la propia complejidad del derecho que dificultan la estricta correspondencia entre los sesgos algorítmicos y la discriminación en sentido jurídico, más centrada en el resultado que en el proceso, donde radica el sesgo algorítmico[14] y, a la vez, "agudizan las tensiones ya existentes en el corpus antidiscriminatorio de la UE[15].

---

[9]Chouldechova, A.: "Fair prediction with disparate impact: a study of bias in recidivism prediction instruments", 2016, pp. 1-17, https://arxiv.org/abs/1610.07524.

[10] Cfr. Diccionario panhispánico del español jurídico, https://dpej.rae.es/lema/equidad. Vid. asimismo http://www.enciclopedia-juridica.com/d/equidad/equidad.htm.

[11] Castán Tobeñas, J. *La idea de equidad y su relación con otras ideas morales y jurídicas afines*, Madrid, Reus, 1950.

[12] Ruiz-Gallardón, I.: "La equidad: una justicia más justa". *Foro, Nueva época*, vol. 20, núm. 2, 2017, pp. 175 y 183, http://dx.doi.org/10.5209/FORO.59013.

[13] Hildebrandt, M.: "The issue of bias. The framing powers of ML". *Computer Science*, DOI:10.2139/ssrn.3497597. En M. Pelillo y T. Scantamburlo (eds.): *Machine We Trust. Perspectives on Dependable AI*, MIT Press 2021, http://dx.doi.org/10.2139/ssrn.3497597 (Preprint). P. 13.

[14] Vantin, S.: "Inteligencia artificial y derecho antidiscriminatorio", en Llano Alonso, F., y Garrido Martín, J. (eds.): *Inteligencia artificial y derecho. El jurista ante los retos de la era digital*. Aranzadi, Cizur Menor, 2021, p. 370; Foster, S. R.: "Causation in antidiscrimination law: beyond intent versus

2. Sesgos técnicos frente a discriminación jurídica

El uso de mecanismos automatizados para la adopción de decisiones traslada a algoritmos de decisión las elecciones humanas, lo que comporta siempre la selección de un elemento de entre una categoría y puede incorporar arbitrariedad. La injerencia de terceros en la conformación de tales criterios, junto con la autonomía del aprendizaje automático, distorsiona tanto la forma en que se adoptan las decisiones como el modo de operar los criterios con respecto a los mecanismos "humanos" tradicionales, no exentos de sesgos ni de arbitrariedad. A estos pueden superponerse las propias indicaciones humanas (v.g. de la empresa) para el diseño del algoritmo, o, en su caso, las de los desarrolladores del software, asumiendo en este caso los sesgos de programación o de entrenamiento que hayan podido intoxicar el modelo en cuestión, si es que no se ha entrenado en un contexto diferente a aquel donde deba operar. Habida cuenta de todo ello, los sesgos provenientes de mecanismos automatizados basados en IA, como los que operan con aprendizaje automático y macrodatos, ¿pueden asimilarse al concepto jurídico de "discriminación"?

En el ámbito del trabajo, según el Convenio de la OIT sobre la discriminación (empleo y ocupación), 1958 (núm. 111), la discriminación es «cualquier distinción, exclusión o preferencia basada en motivos de raza, color, sexo, religión, opinión política, ascendencia nacional u origen social que tenga por efecto anular o alterar la igualdad de oportunidades o de trato en el empleo y la ocupación». Si el concepto se corresponde con la diferencia de trato basada en las características personales de un individuo, y no se anuda al propósito de discriminar, sino al resultado[16], el sesgo -no deliberado- procedente del análisis algorítmico puede identificarse, por tanto, con el concepto jurídico de discriminación, en el contexto del citado convenio. Si se contrasta este concepto con el de las directivas de la Unión Europea contra la discriminación, aplicables en el ámbito del trabajo (Directivas 2000/78, 2000/46 y 2006/54), se confirma que, de igual modo, sea cual sea el ámbito de la directiva, aquel se identifica con la situación en la que "una persona sea, haya sido o pudiera ser tratada de manera menos favorable que otra en situación análoga por alguno de los motivos mencionados en el artículo 1" (de la respectiva directiva: en el caso de la 2000/78, "religión o convicciones, de discapacidad, de edad o de orientación sexual en el ámbito del empleo y la ocupación", en el ámbito de la Directiva 2000/43, raza u origen étnico, y en el caso de la Directiva 2006/54, sexo)[17]. De forma que, sin perjuicio de la delimitación del alcance de la responsabilidad por actos discriminatorios, el concepto de discriminación

---

impact". *Houston Law Review*, núm. 41, 2004; y Xenidis, R. y Senden, L.: "EU non-discrimination Law in the era of artificial intelligence: mapping the challenges of algorithmic discrimination", U. Bernitz et al (eds.): *General Principles of EU law and the EU Digital Order*. Kluwer Law International, 2020, p. 172.

[15] Xenidis, R. y Senden, L.: "EU non-discrimination Law in...", op. cit., p. 738.

[16] Tomei, M.: "Análisis de los conceptos de discriminación y de igualdad en el trabajo". *Revista Internacional del Trabajo*, vol. 122, núm. 4, 2003.

[17] El anteproyecto de Ley para la igualdad real y efectiva de las personas trans y para la garantía de los derechos de las personas LGTBI (https://www.igualdad.gob.es/servicios/participacion/audienciapublica/Documents/APL%20Igualdad%20Trans%20+LGTBI%20v4.pdf) incluye también dichas categorías protegibles, aunque no lo hace dentro de la categoría de "sexo", sino como categoría autónoma.

se identifica con una diferencia de trato basada en una causa protegida[18], *aun cuando esta conducta o acto no sean intencionados o buscado el resultado discriminatorio.*

Este rasgo permite abarcar todos los efectos derivados de sesgos algorítmicos, se trate de sesgos buscados o de sesgos accidentales derivados de la inferencia de datos. Al mismo tiempo, permitiría cubrir las situaciones interseccionales cuando el rasgo relevante para el algoritmo no sea una categoría protegida, sino otro rasgo "secundario" si aparece o se asocia a un rasgo sí cubierto, salvando la falta de cobertura legal de la interseccionalidad como categoría jurídica propia, en tanto pueda colaborar o influir en el resultado sesgado que provoque sobre categorías sí protegidas (v.g. el rasgo determinante para el algoritmo puede ser el sobrepeso, pero, asociado a una mujer, categoría protegida, puede tener relevancia jurídica para informar una potencial reclamación por discriminación basada en el sexo). Lo cual implica conceder un valor jurídico divergente del resultante del sesgo algorítmico, porque, para el sistema de IA, el valor específico estriba en la confluencia de las inferencias (en este ejemplo, sexo + aspecto), mientras que una sola de las causas probablemente habría producido un resultado distinto, que pudiera no ser determinante para justificar un indicio suficiente a fin de plantear una reclamación por discriminación. A título de ejemplo, inferir del código postal o lugar de residencia, o del nombre u otra variable la pertenencia a categorías protegidas implica combinar distintos factores en el análisis, que, por separado, podrían ser "inocuos" o irrelevantes, pero que, actuando conjuntamente, confluyen en un resultado que afecta, en el ejemplo considerado, a personas de cierto origen nacional o étnico, religión o nivel económico, de suerte que esta conclusión sea la que determine el posicionamiento del individuo o individuos en un determinado ranking, orden, o situación, con efectos perjudiciales en el ámbito laboral (v.g. no seleccionado para ocupar un puesto de trabajo). Por otra parte, acreditar el indicio de discriminación se torna especialmente difícil, en tanto que debe consistir en hallar la correlación entre los datos, que puede ser una cuestión de "modelo de caja negra" (propio del aprendizaje profundo, donde se desconocen las razones que motivan el resultado a partir de la introducción de macrodatos[19]), lo que significa que la IA multiplica el sesgo, pero no contribuye a identificarlo, al basarse, cuando se emplea aprendizaje automático, en correlaciones de datos.

Los datos son, en consecuencia, el peor obstáculo a la aportación de indicios, sin perjuicio de que siga siendo posible emplear los mecanismos tradicionales al efecto, esto es, la existencia de elementos protegidos en el caso y el resultado negativo para quien los alega. Del mismo modo, el aprendizaje automático permite mayores justificaciones para desvirtuar los indicios[20], toda vez que opera en el terreno de la discriminación indirecta, por tanto, desarticulable por justificación objetiva y razonable. Por otra parte, la víctima será en estos casos menos consciente del sesgo[21], principal barrera en muchos casos para la visibilización del riesgo y de su impacto.

---

[18] Art. 21 de la Carta de los Derechos Fundamentales de la Unión Europea: por razón de sexo, raza, color, orígenes étnicos, religión o convicciones, y art. 14 CE: por nacimiento, raza, sexo, religión, opinión o cualquier otra condición o circunstancia personal o social.

[19] Mayson, S. G.: "Bias In, Bias Out". *The Yale Law Journal*, vol. 128, núm. 8, 2019.

[20] Xenidis, R. y Senden, L.: "EU non-discrimination law in the era...", op. cit., p. 747.

[21] Ebers, M.: "Ethical and legal challenges", en Ebers, M. y Navas, S., dirs.: *Algorithms and law.* Cambridge University Press, 2020, p. 79.

B. Insuficiencia del derecho antidiscriminatorio

1. Discriminación múltiple e interseccionalidad

La discriminación *múltiple* forma parte de estudios[22], de textos programáticos y de iniciativas legislativas de la Unión Europea, conscientes de la prevalencia del género en este fenómeno multicausal[23], pero no ha sido efectivamente integrada en su *corpus iuris*, lo que impide su efectiva aplicación, pese a su confluencia necesaria con el principio de transversalidad de género previsto en el art. 4 de la Ley Orgánica 3/2007, de 22 de marzo, para la igualdad efectiva de mujeres y hombres[24]. No así la discriminación interseccional como forma específica de discriminación donde prevalece el resultado final sobre la suma de causas o multiplicidad de estas[25].

La discriminación *interseccional*, término acuñado por Crenshaw[26], parte de la necesidad de considerar de manera específica las situaciones de discriminación que operan por la confluencia de distintos factores y que provocan un resultado distinto de la mera suma de causas independientes. En el plano digital, el sistema de inferencias que realiza el aprendizaje automático puede rastrear distintos factores, no todos ellos objeto de tutela específica bajo la norma constitucional o la legalidad de cada país (caso de la obesidad, el aspecto físico, el uso de determinada indumentaria, tatuajes, piercings en España… fenómeno denominado *profiling* o "aspectismo", actuación basada en el aspecto, que incorpora con frecuencia prejuicios de género), cuya interrelación provoque el resultado final, sea la no priorización en un proceso de selección, sea la evaluación negativa a otros efectos… El análisis interseccional, sostiene Serra Cristóbal[27], "permite analizar las interdependencias entre diversos factores de opresión y, de manera simultánea, promover una interpretación indivisible e interdependiente de los derechos humanos"[28].

La falta de positivización hasta ahora de la figura, tanto en el derecho de la Unión Europea (donde se alude en el apartado 14 de la Directiva 2000/43, y en el apartado 3 de la Directiva 2000/78, ambos en el preámbulo y sin contenido concreto anudado en el cuerpo de la norma)[29], como en el español (sin perjuicio de que la Proposición de Ley

---

[22] Comisión Europea: *Tackling Multiple Discrimination. Practices, policies and laws*, 2007, y *Multiple discrimination in EU Law. Oportunities for legal responses to intersectional gender discrimination*, 2009.

[23] Serra Cristóbal, R.: "El reconocimiento de la discriminación múltiple por los tribunales". *Teoría y derecho*, núm. 27, 2020, p. 146. DOI: https://doi.org/10.36151/td.2020.008.

[24] Barrère Unzueta, M. A.: "La interseccionalidad como desafío al mainstreaming de género en las políticas públicas", *Revista Vasca de Administraciones Públicas*, núm. 87-88, 2010, pp. 225-227.

[25] Makkonen, T.: *Multiple, compound and iintersectional discrimination: bringing the experiences of the most marginalized to the fore.* Institute For Human Rights, Abo Akademi University, 2002.

[26] Crenshaw, K.: "Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics". *The University of Chicago Legal Forum: 1989*. HeinOnline--U. Chi. Legal F. 139 1989, https://philpapers.org/archive/CREDTI.pdf?ncid=txtlnkusaolp00000603.

[27] Serra Cristóbal, R.: "El reconocimiento de la discriminación múltiple …", op. cit., p. 157.

[28] Vid. Schiek, D. y Lawson, A. (dirs.): *European Union Non-Discrimination Law and Intersectionality: Investigating the triangle of racial, gender and disability discrimination*, Londres-Nueva York: Routledge, 2016, y Serra Cristóbal, R. (coord.): *La discriminación múltiple en los ordenamientos jurídicos español y europeo*, Valencia: Tirant lo Blanch, 2013.

[29] Para Xenidis y Senden (op. cit., p. 738), el concepto de discriminación múltiple sí está presente en la Directiva 2006/54. En el ámbito europeo, la Sentencia del Tribunal Europeo de Derechos Humanos, asunto B.S. vs. España, de 24 de julio de 2012, sí aplica el concepto aludido. Por otra parte, la propuesta

Integral de Igualdad de 2021 del Grupo Socialista sí la incluya entre sus conceptos en su art. 6.3[30]) ha dejado sin mecanismos jurídicos la tutela efectiva de las situaciones definidas por la confluencia de varias causas de discriminación, como demuestra la STJUE de 24 de noviembre de 2016, asunto C-443/15, *Parris*. De suerte que la interrelación entre los conceptos computacionales y jurídico carece, en la práctica, de efectos concretos, mientras no encuentre respaldo legal, pese a que el impacto del sesgo algorítmico pueda ser mucho mayor, en tanto que las inferencias que puede realizar el aprendizaje automático son capaces de detectar rasgos que a los humanos y sus prejuicios sociales les pasarían inadvertidos, evitando su impacto discriminatorio, lo que determina que esta modalidad de discriminación encierre un potencial mucho más grave que la discriminación simple o monocausal[31], del mismo modo que el aprendizaje automático tiene una más elevada capacidad de generar discriminaciones por asociación, precisamente por su capacidad de inferir correlaciones entre datos[32]. Este potencial discriminatorio se endurece todavía más si se considera su invisibilidad ante las herramientas de control antidiscriminatorias, frente a las cuales puede permanecer oculto o inadvertido. En consecuencia, reorientar el derecho antidiscriminatorio de la UE contribuiría, sin duda, a mitigar los sesgos algorítmicos, lo que intensifica la urgencia de gestionar una cuestión que lleva demasiado tiempo pendiente de resolverse en el ámbito europeo y, por ende, en el español. Wachter[33] propone una simple operación de incluir como categorías protegidas a aquellos individuos que se relacionan con las protegidas de una forma clara por parte del derecho.

En el plano procesal, por tanto, la admisión plena de la figura reforzaría la tutela para las víctimas de discriminación múltiple e interseccional, más allá de haber de aportar indicios de cada una de las causas concurrentes, simplificando la aportación de un solo indicio del sesgo interseccional, también en el ámbito de la discriminación provocada por el uso de sistemas de IA. En esta línea, se ha afirmado que la admisión de la discriminación interseccional como figura jurídica en el derecho antidiscriminatorio facilitaría la actividad probatoria en torno al sesgo algorítmico[34].

---

de directiva horizontal de 2008, aún no aprobada, 11531/08 SOC 411 JAI 368 MI 246 - COM(2008) 426 final (texto consolidado de 2017), alude expresamente a la interseccionalidad y composición de causas de discriminación de entre las protegidas por las directivas, prohibiendo en su art. 2.2 a) la discriminación múltiple.

[30] Dicho precepto establece que "se produce discriminación múltiple cuando una persona es discriminada de manera simultánea o consecutiva por dos o más causas de las previstas en esta Ley" (a) y "se produce discriminación interseccional cuando concurren o interactúan diversas causas de las previstas en esta Ley, generando una forma específica de discriminación" (b). Asimismo, dispone que "en supuestos de discriminación múltiple e interseccional la motivación de la diferencia de trato, en los términos del apartado segundo del artículo 4, debe darse en relación con cada uno de los motivos de discriminación" (en https://www.congreso.es/public_oficiales/L14/CONG/BOCG/B/BOCG-14-B-146-1.PDF). También el proyecto de Ley Orgánica de Garantía Integral de la Libertad Sexual (aprobado por el Consejo de Ministros el 6/7/2021, https://transparencia.gob.es/servicios-buscador/contenido/normaelaboracion.htm?id=NormaEV03L0-20200902&lang=es&fcAct=2021-06-30T12:23:27.739Z) introduce en su art. 2.5 el principio de atención a la discriminación interseccional y múltiple, que define como la superposición a la violencia de otros factores de discriminación.

[31] Xenidis, R. y Senden, L.: "EU Non-Discrimination Law in the Era of Artificial Intelligence...", op. cit., p. 740.

[32] Wachter, S.: "Affinity profiling and discrimination by association in online behavioural advertising". *Berkeley Technology Law Journal*, núm. 35, 2020, p. 371.

[33] Wachter, S.: "Affinity profiling and discrimination by association …", op. cit., p. 373.

[34] Vantin, S.: "Inteligencia artificial y derecho antidiscriminatorio", op. cit., p. 276.

2. Equivalencias entre sesgos algorítmicos y discriminación por asociación, por error o múltiple

Dada la realidad contrastada del impacto discriminatorio del uso de IA, conviene constatar si nuestro derecho está preparado para dar respuesta satisfactoria a la protección del derecho a la igualdad y no discriminación cuando se plantea una reclamación sobre responsabilidad algorítmica. Y ello requiere analizar si el derecho vigente da respuesta a las formas de discriminación causadas por el uso de aprendizaje automático y, en particular, inferencias de datos.

Como señalan Gerards y Xanidis[35], la sustitución de criterios asociados a características protegidas por el derecho antidiscriminatorio de la UE no es admitida por su Tribunal de Justicia (v.g. país de nacimiento como sustituto de origen étnico, como sucedió en la sentencia de 6 de abril de 2017, asunto C-668/15, Jyske Finans A/S[36]), pese a la efectiva existencia de lo que Xenidis denomina formas interseccionales de discriminación que cuestionan la gramática del derecho antidiscriminatorio de la UE[37]). Esta postura, poco coherente con el impacto discriminatorio de la IA, permite concluir que el derecho de la UE (y, consiguientemente, el español), no están preparados para dar respuesta a la discriminación algorítmica, pues permiten a distintas manifestaciones del impacto de las decisiones automatizadas escapar de un marco legal encorsetado que adolece de vacíos legales y obliga a constreñir el análisis del ámbito o alcance de la discriminación a unos parámetros simplificados, donde no hay espacio para la discriminación múltiple[38], un concepto en sí mismo complejo y problemático[39], que asoma tímidamente en los más recientes textos de la UE, como la Estrategia de Igualdad de Género 2020-2015 (se refiere a la interseccionalidad) o la vieja propuesta de directiva horizontal de 2008 (que en su texto de 2017 incorporaba la prohibición de discriminación múltiple), así como en la doctrina del TJUE (pese a no materializarse en sus resoluciones finalmente, como demuestran los casos Meister y Parris).

Las decisiones automatizadas pueden incurrir en discriminación múltiple como consecuencia de su reproducción de la realidad social compleja y donde convergen por intersección distintos prejuicios sociales susceptibles de confluir sobre los mismos individuos[40]. Identificar y corregir este sesgo cuando proviene de modelos matemáticos para la adopción de decisiones aumenta la complejidad de una tarea *per se* críptica, pues, en términos computacionales, exige la combinación de diversos criterios de

---

[35] Gerards, J. y Xenidis, R.: "Algorithmic discrimination in Europe: Challenges and Opportunities for EU equality law", *European Futures*, 3/12/2020, https://www.europeanfutures.ed.ac.uk/algorithmic-discrimination-in-europe-challenges-and-opportunities-for-eu-equality-law/. Sobre este mismo tema, vid. Gonzalo Quiroga, M.: "Discriminación racial y control de identificación policial: valoración de la raza como indicio de extranjería y de nacionalidad", *La Ley: Revista jurídica española de doctrina, jurisprudencia y bibliografía*, núm. 3, 2001, pp. 2158-2164.

[36] Gerards, J. y Xenidis, R., op.cit en nota anterior.

[37] Xenidis, R.: "Tuning EU equality law to algorithmic discrimination: Three pathways to resilience", *Maastricht Journal of European and Comparative Law* 2020, Vol. 27, núm. 6, 4/1/2021, pp. 736-758, https://doi.org/10.1177/1023263X20982173.

[38] FRA – Agencia de los Derechos Fundamentales de la Unión Europea: *Inequalities and multiple discrimination in access to and quality of healthcare*, 2010, http://fra.europa.eu/en/publication/2013/inequalities-discrimination-healthcare.

[39] Rey Martínez, F.: "La discriminación múltiple, una realidad antigua, un concepto nuevo". *Revista española de derecho constitucional*, año 28, 2008, núm. 84, p. 3.

[40] Xenidis, R.: "Tuning EU equality law to algorithmic discrimination: …, cit., p. 739-740.

corrección para evitar la segregación de una sola de las características a tutelar. Pero, en realidad, la técnica pone al alcance del derecho antidiscriminatorio opciones de mucha mayor complejidad jurídica, pues resulta más accesible diseñar una orden de no discriminar (esto es, de priorizar) determinadas características combinadas que confluyen en un mismo individuo (sexo, origen ético, orientación sexual…) para el algoritmo que responder con las herramientas legales necesarias a un caso de discriminación múltiple. Simplemente porque lo que, para la técnica resulta factible, para el derecho no está aún consolidado, pues el derecho de la UE[41], como el español, no han positivizado aún un concepto legal de discriminación múltiple que permita identificar y tutelar este tipo de intersección discriminatoria, en lugar de mantener su tratamiento como suma de causas independientes[42].

Por otra parte, el concepto de *discriminación por asociación* (concepto recogido explícitamente en el art. 2 e) del Texto Refundido de la Ley General de Derechos de las Personas con Discapacidad y de su Inclusión Social, LGDPDIS (Real Decreto Legislativo 1/2013, de 29 de noviembre) ofrece dudas aun irresolutas. En efecto, en tanto no cabe identificar asociación con correlación de datos de *proxy*, pues este solo es un elemento que detecta características por proximidad, pero que no opera por sí mismo como elemento de discriminación, no resulta posible sostener un concepto "digital" de discriminación por asociación. Pero sí cabe afirmar una clara conexión entre ambos conceptos y un elevado potencial discriminatorio derivado de la mayor precisión que la asociación entre la persona objeto de discriminación y aquella característica presente en otras personas próximas a su círculo que pueda rastrearse por un sistema de IA (v.g. relación con una persona con discapacidad). Pues las correlaciones que pueden ser desconocidas por los humanos (pero que, conocidas, pudieran llevar a discriminar) son más fácilmente captables por un sistema automatizado (por inferencias en conjuntos de datos) y, por ende, pueden amplificar la capacidad asociativa con fines peyorativos (el ejemplo más claro, que proporciona la conocida sentencia Coleman, de 17 de julio de 2008 [asunto C-303/06], es la detección algorítmica de la conexión con una persona con discapacidad -familiar directo- a través de distintas inferencias, por ejemplo, su afiliación a una asociación de progenitores de personas con diversidad funcional, u otro tipo de datos que evidencien que, pese a que la empresa desconocía este extremo de su vida familiar, existe tal asociación, como rasgo tributario de exclusión en el proceso de selección y contratación o a efectos de otras decisiones empresariales).

En realidad, la correlación de inferencias entre datos implica relacionar entre sí rasgos que el algoritmo o conjunto de algoritmos empleados priorizan o aquellos que descartan, en función del resultado óptimo que proporcione su análisis, lo que acaba implicando que ciertos rasgos o características sean definidos como más adecuados para

[41] La STJUE de 24 de noviembre de 2016, asunto C-443/15, *Parris*, rechaza la consideración como discriminación múltiple o interseccional de dos causas si por separado no son discriminatorias, es decir, exige que por separado ambas sean discriminatorias, lo que, en realidad, podría calificarse de superposición de causas, pero no de una interseccionalidad con mayor carga peyorativa asociada a la conjunción de causas en un resultado nuevo. A tenor de la Proposición de Ley Integral de Igualdad promovida por el grupo parlamentario socialista, la discriminación interseccional se define por la concurrencia o interactuación de diversas causas protegidas, generando una forma específica de discriminación (art. 6.3 b). Por otra parte, dicho texto distingue de este concepto el de discriminación múltiple, para asimilarlo al recogido por la STJUE citada, como la situación en la que "una persona es discriminada de manera simultánea o consecutiva por dos o más causas" protegidas (art. 6.3 a).

[42] Xenidis, R.: "Tuning EU equality law to algorithmic discrimination…", op. cit., p. 741.

los fines empresariales pretendidos (v.g. empleo), mientras que otros serán relegados. Sea como sea, en esa operación automatizada será altamente factible detectar rasgos que determinen una discriminación *por asociación*, en el sentido anteriormente indicado, y explicitado legalmente en el art. 2 e) LGDPDIS, en el derecho de la UE, y en el texto de la Proposición de Ley integral de igualdad aludido (2021), en el que se predica respecto de todas las causas en ella protegidas (más allá de la discapacidad) y se describe como la situación en la que "una persona o grupo en que se integra, debido a su relación con otra sobre la que concurra alguna de las causas [protegidas] es objeto de un trato discriminatorio" (art. 6.2 a).

La *discriminación por error* (aquella que se funda en una apreciación incorrecta acerca de las características de la persona o personas discriminadas) parece identificarse *a priori* con los errores del aprendizaje automático al realizar inferencias que atribuyen a ciertos rasgos consecuencias que pueden no resultar concordantes con la realidad (errores derivados de las reglas de la lógica cuando existe una insuficiente base de datos, sobrerrepresentación o infrarrepresentación de categorías de datos que a su vez están codificando el mundo real). El proceso deductivo mental que conduce a apreciaciones erróneas puede ser humano o computacional, aunque en buena medida siempre cabe hallarse el error humano en el segundo caso, si se considera que el etiquetado de datos para transformar elementos de la realidad en conocimiento digital también es realizado por humanos, que pueden incurrir en errores y sesgos.

Ahora bien, la cuestión central en este caso es la autoría de tal error, pues, si en el caso humano es claramente atribuible a su incorrecta comprensión de la realidad o de un rasgo concreto de la persona objeto de valoración u observación, en el caso digital este error ha sido cometido por otras personas de forma indirecta o bien por el propio mecanismo de aprendizaje automático, en definitiva responsabilidad también de personas[43]. En el primer caso la imputación de responsabilidad es subjetiva, mientras que en el segundo puede ser objetiva si, en el plano jurídico-laboral, se atribuye la responsabilidad a quien emplea las herramientas que conducen al sesgo, si, al fin y al cabo, los empleadores son los sujetos responsables a todos los efectos en su relación jurídica frente a los empleados o "empleables" *ex* arts. 1902, 1911 C.c., y se deriva de los arts. 42 y 44 ET cuando regula la responsabilidad empresarial en caso de intervención de terceras empresas (subcontratación y transmisión de empresa) o en materia de uso de productos defectuosos en el marco de los riesgos laborales (art. 41 LPRL). El modelo de responsabilidad laboral de carácter contractual salva, en estos casos, las dificultades propias de la opacidad y difícil trazabilidad de sujetos responsables en toda la cadena computacional que ha conducido al resultado final, ya sea por el diseño, por la alimentación, por el etiquetado de datos, el entrenamiento en contextos distintos al de su creación, o la aplicación, así como la de identificar a un responsable principal o único[44].

[43] Así lo entiende también el Parlamento Europeo cuando en el apdo. 7 de la Resolución de 20/10/2020, con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial (2020/2014(INL), donde "observa que todas las actividades, dispositivos o procesos físicos o virtuales gobernados por sistemas de IA pueden ser técnicamente la causa directa o indirecta de un daño o un perjuicio, pero casi siempre son el resultado de que alguien ha construido o desplegado los sistemas o interferido en ellos".

[44] Vantin, S.: "Inteligencia artificial y derecho antidiscriminatorio", op. cit., p. 376.

C. Sobre la valoración como discriminatorios de las decisiones automatizadas

1. ¿Es relevante identificar el impacto de las arquitecturas computacionales a efectos de calificarlo como discriminatorio jurídicamente?

La pregunta que se formula tiene una directa conexión con un abordaje práctico de la cuestión, desde la perspectiva del análisis de las responsabilidades empresariales derivadas del resultado discriminatorio. Así pues, la pregunta es si es realmente necesario saber cómo ha llegado a una conclusión un sistema de IA, o si debemos atenernos en exclusiva al resultado.

Por tanto, en este plano teórico-disyuntivo, podríamos barajar dos hipótesis: a) la *tesis de la conducta*; b) la *tesis del resultado*. Nuestro derecho prima el resultado, pero también concede relevancia a la conducta en sí, porque esta permite determinar el alcance de la gravedad del incumplimiento como de la responsabilidad derivada, aun cuando la mera existencia de un impacto discriminatorio constituye el indicio necesario para desplazar, en un plano procesal, la carga probatoria hacia el autor o autores de la conducta que desencadena el resultado hipotéticamente discriminatorio. En consecuencia, no es irrelevante el proceso que conduce al resultado ni los factores que intervienen en este, y, por otra parte, resulta de indudable interés para conocer ambos extremos determinar cómo operan tales sesgos automatizados y cómo calificarlos desde la perspectiva del derecho.

Lo que sí parece imprescindible es que los expertos en computación e IA dominen la terminología jurídica afectada para tratar no solo de hallar las correspondencias entre ambos lenguajes, sino también para abordar la mitigación de sesgos con el bagaje adecuado. A título de ejemplo, conocer por los científicos de datos la prohibición legal de la discriminación por asociación contribuiría a entender el impacto que podrían tener las inferencias de datos sobre el círculo personal próximo de cada sujeto acerca del trato que finalmente se le dispensará, porque estas correlaciones no son ni irrelevantes ni neutrales para el derecho. Por ello, se abordarán seguidamente tales correspondencias, en aquello que resulte de especial interés para la identificación y calificación de conductas y decisiones.

2. Impacto discriminatorio de los algoritmos

Más allá de la regulación de la dinámica que los algoritmos provocan sobre la esfera de protección de los derechos de los trabajadores, resulta imprescindible su evaluación desde la perspectiva del derecho a la igualdad y a la no discriminación, así como dilucidar si nuestro derecho está preparado para dar respuesta satisfactoria a las reclamaciones sobre responsabilidad por decisiones automatizadas. Este planteamiento supone dos tipos de análisis: en primer lugar, la naturaleza de estas formas discriminación a la luz de nuestro derecho y en particular de las directivas antidiscriminación de la UE (Directivas 2000/78, 2000/43 y 2006/54), todas ellas aplicables en el ámbito del empleo, y, en segundo lugar, su posible calificación como discriminación directa o indirecta.

El análisis del impacto discriminatorio de las decisiones automatizadas obliga a confrontarlas con las reglas del derecho antidiscriminatorio, que lleva a formular diversas cuestiones, algunas de las cuales exceden el análisis del prisma discriminatorio: a) ¿puede haber intencionalidad discriminatoria en el uso de un algoritmo que ofrece conclusiones sesgadas?; b) ¿es necesario que la haya o cabe imputar responsabilidad

objetiva derivada de la interposición de mecanismos automatizados para adoptar decisiones o bien cabe deducir la nulidad de estas por falta de intervención humana?; c) basar decisiones en técnicas de automatización con escasa intervención humana que provocan sesgos discriminatorios ¿constituye discriminación directa o discriminación indirecta? Finalmente, procede verificar si las herramientas de tutela tradicionales son suficientes o conviene una reinterpretación de estas formas de discriminación, en tanto fueron diseñadas para operar con otros parámetros.

En el marco del derecho antidiscriminatorio de la UE, del que resulta tributario el español, y siguiendo a Miné[45], "la discriminación directa puede ser intencional y explícita con respecto al motivo prohibido", "pero, al estar dicha discriminación explícitamente afirmada, en especial en una norma, cada vez con menor frecuencia, el derecho pone el énfasis en el efecto producido por la diferencia de trato, según un concepto objetivo de la discriminación", y "el carácter intencional de la discriminación ya no constituye un elemento esencial".

En el ámbito laboral, tampoco alcanza tal relevancia que el propósito económico haya primado, aun sin especial finalidad discriminatoria, este interés sobre la protección de los derechos fundamentales, o si realmente el objetivo buscado es tanto el de hacer de peor condición a unos colectivos o individuos con características compartidas sobre otros que corresponden a un patrón estándar, como el resultado, la discriminación. Como tampoco es obstáculo a tal calificación que el individuo afectado por el trato discriminatorio lo sea por su condición de proximidad con el que pertenece al colectivo protegido (discriminación *por asociación*). Lo relevante es la diferencia de trato entre dos situaciones comparables. Obviamente, estas situaciones pueden estar condicionadas por el recurso a herramientas técnicas basadas en IA o, por el contrario, no haberse basado en ambos casos en los mismos recursos técnicos. No obstante, los elementos a comparar son las situaciones en sí mismas y no cómo se actuó sobre ellas (en este caso, cómo se tomó la decisión), por lo que estos otros aspectos deberán en todo pesar en la valoración del comportamiento hipotéticamente discriminatorio. Pero lo que también *altera las reglas del juego* es que la supuesta anonimización de los datos objeto de tratamiento o alimentación del algoritmo no garantiza la igualdad, pues permite seguir realizando inferencias de otro tipo de datos de *proxy*.

Por lo que interesa a la cuestión analizada, la decisión automatizada puede responder tanto al concepto de discriminación directa como al de discriminación indirecta[46], pues el algoritmo puede estar diseñado con finalidad estricta de descartar ciertas características, por ejemplo, cuando se trata de selección de personal o para realizar una evaluación a partir de criterios aparentemente neutros que perjudiquen a los individuos

---

[45] Miné, M.: "Lucha contra la discriminación: Las nuevas directivas de 2000 sobre la igualdad de trato", ponencia de 31/3-1/4/2003, Trier, p. 5, en http://www.era-comm.eu/oldoku/Adiskri/02_Key_concepts/2003_Mine_ES.pdf.

[46] Con arreglo al art. 6 de la Ley Orgánica 3/2007 (como para las personas por discapacidad, según el art. 2 LGDPDIS), "se considera discriminación directa por razón de x la situación en que se encuentra una persona que sea, haya sido o pudiera ser tratada, en atención a su x, de manera menos favorable que otra en situación comparable", y "discriminación indirecta por razón de x la situación en que una disposición, criterio o práctica aparentemente neutros pone a personas de un x en desventaja particular con respecto a personas del otro, salvo que dicha disposición, criterio o práctica puedan justificarse objetivamente en atención a una finalidad legítima y que los medios para alcanzar dicha finalidad sean necesarios y adecuados" (sustitúyase x por sexo o por discapacidad).

con ciertas características o que deliberadamente las ignoren cuando son claramente condicionantes de la valoración de su productividad (v.g. enfermedad o discapacidad), lo que podría calificarse como "diseño no inclusivo" del algoritmo (protegible, por tanto, desde la perspectiva de la LGDPDIS). Tanto si se trata de una práctica no neutral como de un uso aparentemente neutro, pero susceptible de implicar una desventaja particular para las personas que respondan a uno o más criterios (o bien supondrían una desventaja particularmente para personas en función del sexo, en relación con las personas del otro sexo), lo cierto es que la decisión empresarial resulta constitutiva de discriminación. A menos que pueda operar la salvedad que desvirtúa la presunción de discriminación indirecta, eso es, que el criterio o práctica sean justificados objetivamente por un objetivo legítimo y los medios usados sean apropiados y necesarios. En definitiva, el derecho antidiscriminatorio necesita de una reinterpretación a la luz de estas nuevas formas de discriminación tecnológica que permitan ajustar las respuestas legales.

Admitido que las decisiones automatizadas pueden ser, como las puramente humanas, susceptibles de generar trato discriminatorio injustificado, habrá de valorarse si este es calificable como discriminación directa o como discriminación indirecta.

3. Calificación como discriminación directa o indirecta

Calibrar el impacto discriminatorio de las decisiones basadas en algoritmos obliga a examinar en primer lugar si realmente la empresa traslada o introduce parámetros de sesgo en el algoritmo cuando adopta tales decisiones. La respuesta no es unívoca, porque el propio diseño del algoritmo se construye sobre la base de órdenes que persiguen un objetivo, y este se define por quienes ordenan su programación, sean estos los empresarios a cuyo uso directo se destina, o sean los comercializadores de software de empresa para su adquisición por empleadores con el fin de organizar sus "recursos humanos". Por tanto, es posible que la respuesta sea negativa (ello constituiría una explícita discriminación directa, insertada en la estructura del algoritmo).

Ahora bien, si un algoritmo de aprendizaje automático funciona como una caja negra, y simplemente valora todos los elementos en juego para alcanzar la decisión "más sabia" y acertada, la decisión final (sesgada) podría estar contaminada y constituir un supuesto de discriminación indirecta, pero ajena a las intenciones del empleador o la empleadora que lo utilizaron para fundar una decisión. ¿Es relevante, pues, la propiedad del algoritmo para derivar responsabilidad por discriminación? Si partimos de que la intencionalidad es irrelevante para deducir tal calificación como discriminatorio, ¿qué relevancia tendría usar deliberadamente arquitecturas de sesgo respecto de la mera adquisición de *software* que provoca el mismo resultado no buscado? De igual modo, si el modelo de funcionamiento del algoritmo empleado, v.g. en la selección de personal, toma como referencia un patrón histórico, esto es, los antecedentes de sesgo de la propia empresa, ¿estaríamos ante un posible caso de *discriminación directa inconsciente*? La clave la proporciona la actitud de la empresa frente a este riesgo: su pasividad ante el posible efecto perverso de la elección es lo que le acaba convirtiendo en cómplice del algoritmo. Y, por supuesto, las acciones previas que alimentaron y entrenaron al algoritmo y que este reproduce como modelo ideal a seguir, que fueron en su momento ejecutadas por la empresa, porque en tal caso, si bien no resulta responsable de una orden directa de discriminar (esta sería el diseño ex professo con tal fin), sí es responsable de sus acciones pasadas. La cuestión es de enorme interés desde la perspectiva del derecho antidiscriminatorio, en tanto permite considerar la

responsabilidad por comportamientos pasados cuando estos constituyen la base inconsciente de nuevas decisiones tomadas por un tercero (el algoritmo) pero asumidas por los mismos sujetos, como consecuencia de que un algoritmo ajeno a su esfera de decisión ha determinado que deben ser reproducidos para asistir nuevas decisiones, si la empresa desconoce el funcionamiento del aprendizaje automático y sus consecuencias (precisamente por falta de motivación del algoritmo de los elementos en los que se basa para llegar a la conclusión). En este caso, lo verdaderamente relevante será su consentimiento para validar el sesgo que el algoritmo reproduce (confirmar la propuesta del modelo matemático), por lo que, en el plano laboral, el empleador continúa siendo responsable de su acción discriminatoria, e incluso cabe valorar su conducta como constitutiva de discriminación directa y no indirecta (pues la conducta constitutiva de discriminación directa puede basarse en un mecanismo implícito pero consciente, mientras que la discriminación indirecta requiere que el impacto discriminatorio derive de su afectación prioritaria a un colectivo definido por una característica protegida). De igual modo, el algoritmo puede replicar tanto discriminaciones directas como indirectas pasadas (*patrón histórico*). En cualquier caso, lo que sí parece ser obvia es la inaplicación del plan de igualdad de la empresa como consecuencia del empleo de patrones históricos para tomar decisiones.

El análisis jurídico de la correlación entre los mecanismos reputacionales de empresa en la valoración de trabajadores y las decisiones empresariales tiene, también, una importancia central en esta aproximación, en cuanto es de singular relevancia en la arquitectura de los algoritmos empleados la ausencia de parámetros de orden personal, como la salud, la discapacidad, la conciliación de la vida familiar y laboral, o los derechos sindicales, así como para la corrección del impacto de un sistema de evaluación de estas características. Lo cierto es que los modelos algorítmicos de evaluación del rendimiento están ignorando sistemáticamente criterios de orden jurídico-laboral objeto de especial tutela jurídica frente a la discriminación, como es el caso de la discapacidad o la necesidad de conciliar la vida personal y familiar. La cuestión se puede ejemplificar en la sentencia del Tribunal de Bolonia de 31 diciembre de 2020, núm. 29491, que sitúa este rasgo discriminatorio precisamente en su diseño no inclusivo o huérfano de cualquier factor de corrección de situaciones que servirían para justificar por parte de los repartidores de Glovo la cancelación de su disponibilidad para atender un servicio, regida por el algoritmo *Frank*, lo que motiva que este descabalgue del orden de prioridad de llamamiento a trabajadores que se encuentran en tales circunstancias -legalmente justificadas-, y, consecuentemente, por reiteración en el tiempo, ponga en riesgo incluso la propia pervivencia de su puesto de trabajo.

III. TUTELA DESDE EL DERECHO ANTIDISCRIMINATORIO

A. Marcos regulatorios frente a la opacidad de las decisiones automatizadas

Si bien las decisiones empresariales generan responsabilidad directa de los empleadores con independencia de cómo se hayan construido o asesorado aquellas, la ocupación del espacio de decisión laboral por algoritmos altera los tradicionales mecanismos de respuesta, al modificar la capacidad defensiva de los trabajadores, porque sus características amparan la supuesta objetividad de tales decisiones y los presentan como opacos e incontestables. En efecto, el problema de la opacidad es una constante en el

ámbito de las decisiones, empresariales o públicas[47]. El uso masivo de datos y la convergencia de complejidad y apariencia de objetividad ("turbia ilusión de objetividad"[48]) plantean un escenario de confianza y diluyen la percepción de opacidad, pese a las consecuencias tangibles de tales decisiones.

Buena parte del problema de la opacidad se centra tanto en la naturaleza inmaterial y técnica de los algoritmos como en su propia configuración jurídica en tanto que se encuentran sujetos a derechos de propiedad intelectual. Ambos elementos permiten a sus usuarios directos mantener una difusa aura de opacidad sobre sus decisiones cuando estas son automatizadas. Frente a esta barrera de la opacidad actúan los mecanismos de tutela que se derivan del Reglamento 2016/679, del Parlamento Europeo y del Consejo, 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* [RGPD], y del Convenio 108, para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal, de 28 de enero de 1981[49], ambos relativos a la adopción de decisiones automatizadas, en tanto no se desarrollen instrumentos autónomos en este campo (no lo será tampoco la futura Ley de IA de la UE, iniciada en abril de 2021). Esta ausencia de marco regulatorio adecuado y suficiente (más allá de la mera protección de los datos personales que brinda el RGPD[50]) es común en el plano nacional, donde la postura del gobierno español consiste en apoyar iniciativas regulatorias europeas.

Por tal razón, el marco jurídico aplicable exige una adecuada combinación entre las normas citadas y el derecho antidiscriminatorio, de suerte que, mientras las primeras permiten delimitar los parámetros de la motivación de las decisiones empresariales (necesario presupuesto para superar la carga indiciaria de las circunstancias e impacto discriminatorio de aquellas en el ámbito procesal), para valorar la validez de tales decisiones de acuerdo con parámetros de transparencia (que incluye su *explicabilidad**, o derecho a una explicación sobre su funcionamiento[51] y sus posibles excepciones, amparadas en derechos de propiedad intelectual y empresarial) y de intervención humana (*ex* art. 22 RGPD), el derecho antidiscriminatorio proporciona las reglas de distribución de la carga probatoria (que en este caso refuerzan la consustancial obligación de motivación de las decisiones cuando estas se adoptan de forma automatizada, como resultado de la aplicación de la normativa relativa a tratamiento de datos), para indagar en la hipotética naturaleza discriminatoria de tales decisiones. A

---

[47] Burrell, J.: "How the machine 'thinks': understanding opacity in machine learning algorithms". Vol. 3, núm. 1, 2016, https://doi.org/10.1177/2053951715622512.

[48] Stoica, A.-A., Riederer, C. y Chaintreau, A.: "Algorithmic glass ceiling in social networks: The effects of social recommendations on network diversity". *Proceedings of the Web Conference* 2018, Lyon. ACM, Nueva York, 2018, pp. 923–932, https://doi.org/10.1145/3178876.3186140.

[49] Vid. Consejo de Europa (2017): *Study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications*, Committee of experts on internet intermediaries, MSI-NET(2016)06 rev6, https://rm.coe.int/study-hr-dimension-of-automated-data-processing-incl-algorithms/168075b94a.

[50] Soriano Arnanz, A.: "Decisiones automatizadas: problemas y soluciones jurídicas. Más allá de la protección de datos". *Revista de Derecho Público: Teoría y Método*, vol. 1, 2021, pp. 85-127, en http://revistasmarcialpons.es/revistaderechopublico/article/download/535/549.

[51] Vid. Goodman, B. y Flaxman, S.: "European Union regulations on algorithmic decision-making and a 'right to explanation'". *AI Magazine*, vol. 38, núm. 3, 2017, DOI 10.1609/aimag.v38i3.2741.

*Explicabilidad es el término empleado por la doctrina especializada para referirse a la condición de explicables de los algoritmos.

este conjunto de normas en juego se suma el marco regulatorio de la propiedad intelectual sobre las herramientas de automatización de decisiones, que puede introducir complejidad adicional en la determinación de obligaciones y responsabilidades, pues comporta asimismo problemas de foro competente y de imputación o reparto de responsabilidades[52], por más que en el ámbito laboral tales responsabilidades hayan de dilucidarse siempre al margen de la relación contractual entre empresas y trabajadores, marco delimitador de la respuesta directa de las primeras frente a los segundos.

B. Acceso y explicabilidad de algoritmos

Los algoritmos son definidos como "secuencia(s) finita(s) de reglas formales (operaciones e instrucciones lógicas) que hacen posible obtener un resultado a partir de la entrada de información", lo que implica que dos elementos son cruciales desde una perspectiva jurídica: la secuencia de instrucciones (el "código fuente" [53]) y la información o datos que este utiliza (las llamadas "librerías" o conjunto de datos, sobre los que, a su vez, debe proyectarse el análisis jurídico acerca de su propiedad y derechos de uso, según su procedencia, y acceso por cualquier impugnante), sobre los que centrar el llamado derecho de explicabilidad y transparencia, equivalente a la motivación de la decisión que ayudan a asistir.

1. Derecho de explicabilidad y acceso al razonamiento subyacente

Tanto el derecho de la UE como el español conectan ambos núcleos de tutela: los datos personales y las decisiones automatizadas a través de la protección de los datos personales. El puente de conexión hacia la tutela frente a la discriminación es precisamente la afectación de las decisiones automatizadas basadas en datos, en cuanto estas emplean perfilación (*ex* art. 22 RGPD) y pueden contener sesgos, de muy difícil entendimiento, como resultado de la propia complejidad del aprendizaje automático en el que se basan[54]. De ahí que uno de los elementos fundamentales para articular la protección frente a la discriminación algorítmica, hasta alcanzar un mayor desarrollo normativo, sea, precisamente, el análisis del tratamiento de datos, aun admitiendo la limitación del ámbito aplicativo y posibilidades que brinda el art. 22 RGPD y, por supuesto, del art. 25 del mismo texto, que apela a herramientas que se han revelado como absolutamente insuficientes en el campo de las inferencias por aprendizaje automático (la seudonimización y otras técnicas a las que alude dicho precepto para anonimizar datos o despojarlos de rasgos personales, puesto que cualquier tipo de rasgo es susceptible de rastreo e inferencia).

---

[52] Vantin, S.: "Inteligencia artificial y derecho antidiscriminatorio", op. cit., p. 371.

[53] El código fuente se puede definir como "el conjunto de líneas de textos, que son las directrices que debe seguir la computadora para realizar dicho programa; por lo que es en el código fuente donde se encuentra escrito el funcionamiento de la computadora", en caracteres alfanuméricos en un lenguaje de programación elegido por programadores (como pueden ser: Basic, C, C++, C#, Java, Perl, Python, PHP). Aplicado a algoritmos, "por código fuente se entiende todo texto legible por un ser humano y redactado en un lenguaje de programación determinado. El objetivo del código fuente es crear normas y disposiciones claras para el ordenador y que este sea capaz de traducirlas a su propio lenguaje" (definición de *Digital Guide Ionos*).

[54] Gunning, D.: "Explainable Artificial Intelligence (XAI)", 2017, https://www.cc.gatech.edu/~alanwags/DLAI2016/(Gunning)%20IJCAI-16%20DLAI%20WS.pdf.

En el ámbito europeo, dos instrumentos legales garantizan el derecho a la protección de datos: el Convenio para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal, convenio 108, y el Reglamento (UE) 2016/679 (RGPD). En el ámbito interno, la norma es la Ley Orgánica 3/2018, de 5 de diciembre, *de Protección de Datos Personales y garantía de los derechos digitales, para la protección del derecho fundamental de las personas físicas a la protección de datos personales* (LOPD), amparado por el artículo 18.4 de la Constitución[55].

Por lo que respecta al derecho de transparencia y explicabilidad, puede materializar, en el contexto de una reclamación por discriminación, la justificación objetiva y razonable que resulta exigible a la empresa autora de la decisión cuestionada y que constituye aplicación de la distribución de la carga probatoria en un proceso por discriminación. No obstante, hacer inteligibles los parámetros y criterios de las decisiones que se toman (*principio de transparencia algorítmica*) resulta difícil en estos casos. La explicabilidad de los algoritmos cumple diversas funciones: ayuda a entender su funcionamiento, tanto para diseñadores y desarrolladores como para personas afectadas por sus efectos, también contribuye a la confiabilidad del sistema que los utiliza (y a su auditoría) y, asimismo, permite construir el argumentario jurídico para desvelar su posible ilegalidad[56], o la de su impacto una vez aplicado en un contexto determinado. Por ello resulta determinante calibrar el grado de explicabilidad y valorar la exploración neutral de los algoritmos por un tercero que permanezca ajeno a los problemas de propiedad intelectual y secreto empresarial que pueda suponer su revelación a efectos de búsqueda de sesgos[57]. En este terreno es clave la herramienta de la auditoría de algoritmos.

En la configuración de este derecho, la primera cuestión relevante es la exclusión de la protección a los datos *no personales*, *ex* art. 9.1 b) del Convenio 108 (que garantiza el derecho individual a la comunicación de los datos procesados de una forma inteligible, toda la información disponible sobre su origen, sobre el periodo de preservación, así como cualquier otra información, con el fin de garantizar la transparencia del tratamiento), que prevé una excepción: los datos personales que no se recopilan de los interesados, caso en el que el responsable está exento de la obligación si el procesamiento implica "esfuerzos desproporcionados". Es posible interpretar que el aprendizaje profundo complica especialmente el cumplimiento de este deber, pudiendo identificarse esta situación con los "esfuerzos desproporcionados" a los que se refiere el art. 8.3.

En segundo lugar, la propia regulación del derecho la restringe a la *contratación en línea* o *procesos totalmente automatizados*. En efecto, los interesados tienen derecho a conocer la motivación del algoritmo cuando este se utiliza para elaborar perfiles (el caso regulado por el art. 22 RGPD y cuyo objetivo es cubrir el ámbito público de procesamiento de datos, con fines de orden público, como ha entendido nuestro tribunal constitucional y como se desprende de la proposición de Ley de IA de la Unión Europea), es decir, al conocimiento del *razonamiento subyacente* en el procesamiento de datos cuando sus resultados les son aplicados (este es también el sentido apuntado

[55] Vid. Castellanos Claramunt, J., y Montero Caro, M. D.: "Perspectiva constitucional de las garantías de aplicación de la inteligencia artificial: la ineludible protección de los derechos fundamentales". *Ius et Scientia*, vol. 6, núm. 2, 2020, pp. 72-82.
[56] Ebers, M.: "Ethical and legal challenges", op. cit., p. 48.
[57] Ebers, M.: "Ethical and legal challenges", op. cit., p. 80.

por el dictamen del CESE cuando se refiere a que el principio de transparencia algorítmica consiste en hacer inteligibles los parámetros y criterios de las decisiones que se toman), y a que esta explicación sea proporcionada por humanos. El considerando 71 del Reglamento predica este derecho de las personas que acceden a *servicios de contratación en red en los que no medie intervención humana alguna,* y mantiene que e*ste tipo de tratamiento incluye la elaboración de perfiles consistente en cualquier forma de tratamiento de los datos personales que evalúe aspectos personales relativos a una persona física* (afiliación sindical, rendimiento en el trabajo, origen étnico o racial, las opiniones políticas, la religión o creencias filosóficas, datos relativos a la salud o datos sobre la vida sexual, o las condenas e infracciones penales o medidas de seguridad conexas, considerando número 75[58]), *en particular para analizar o predecir aspectos relacionados con el rendimiento en el trabajo… en la medida en que produzca efectos jurídicos en él o le afecte significativamente de modo similar.* Literalmente, pues, la norma parece acotar un ámbito restringido del derecho a la explicabilidad, no extensible, por tanto, a decisiones automatizadas en parte del proceso de decisión, aun cuando esta parte sea especialmente relevante en el conjunto de factores que conducen al a decisión final, v.gr. cribado de currículos en un proceso de selección, finalmente "humanizado" al poner al frente de la decisión a humanos, que, como sucedió en el asunto Uber (Tribunal de Ámsterdam, sentencia C/13/692003/HA RK 20-302), puede haberse limitado a convalidar la recomendación del sistema de IA, que, a su vez, es posible que haya empleado un patrón histórico que replique decisiones sesgadas anteriores, o, aunque ello no sea así, pone fin a un procedimiento en el que la fase de cribado de currículos fue íntegramente automatizada, aunque el proceso haya contado con humanos en su fase final. En este ejemplo, si se considera, v.gr., que han sido presentados mil currículos, de los cuales han sido preseleccionados quince para su examen humano, el grueso del sesgo, donde probablemente radique la mayoría de discriminaciones por diversas causas protegidas, habrá quedado incluido en una fase íntegramente automatizada, sobre la cual no se ofrecerá explicación alguna, cuando, como indica el considerando número 71, no se trate de "*servicios de contratación en red en los que no medie intervención humana alguna"*, lo cual puede acontecer porque no sean servicios de contratación en red o porque sí exista alguna intervención humana en la cadena de decisión. De ahí que devenga de singular relevancia la delimitación del concepto, para exigir que se trate de "intervención humana significativa" o "sustancial".

En tercer lugar, el contenido de la explicación se refiere al "razonamiento subyacente", es decir, al propio *mecanismo de razonamiento del algoritmo* (¿el código fuente o su modificación por aprendizaje profundo?), mientras la norma española sólo ampara el derecho a conocer la finalidad del tratamiento, *no cómo este se ejecuta técnicamente*[59].

---

[58] Los "datos sensibles", a tenor del art. 6 del Convenio 108, son categorías especiales de datos protegidos por el citado precepto, que requieren garantías complementarias apropiadas cuando se procesan, especialmente el origen racial o étnico, opiniones políticas, afiliación sindical, creencias religiosas o de otro tipo, vida sexual o salud…de manera autónoma o en combinación con otros datos (*Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data*, 2015).

[59] La legislación española resulta menos garantista en este aspecto que la europea. En efecto, según ha entendido la STC de 22/5/2019, rec. inconst. núm. 1405/2019, "el tratamiento de las categorías especiales de datos personales es uno de los ámbitos en los que de manera expresa el Reglamento General de Protección de Datos ha reconocido a los Estados miembros "margen de maniobra" a la hora de "especificar sus normas", tal como lo califica su considerando 10". Este margen de configuración legislativa, afirma el tribunal constitucional, se extiende tanto a la determinación de las causas

Lo que suscita un problema de vaciado de contenido en el uso de aprendizaje automático, pues el elemento determinante de la decisión serán los datos. Y, aun cuando los datos sirvan para elaborar perfiles, el alcance del derecho continúa limitándose a esta circunstancia (así como al derecho a ser informado de su derecho a oponerse, de darse las circunstancias del art. 22 del Reglamento, por tanto, también dentro de un marco restringido de aplicación). El art. 11.2.2º LOPD se refiere expresamente a los datos obtenidos para la realización de perfiles, reconociendo el derecho de los afectados a ser informados de su derecho a oponerse a la adopción de decisiones individuales automatizadas que produzcan efectos jurídicos sobre ellos o les afecten significativamente de modo similar, cuando concurra este derecho de acuerdo con lo previsto en el art. 22 RGPD, pero no profundiza más allá de esta remisión para dar contenido al derecho de explicabilidad. Especialmente porque el art. 11.2.3º indica que el responsable de los datos obtenidos para el aprendizaje del algoritmo podrá dar cumplimiento al deber de información establecido en el art. 14 RGPD facilitando la información *básica* señalada. Por ello, el derecho a la información básica de quien concurre a un proceso de selección y resulta afectado por un algoritmo que toma la decisión a partir de análisis de datos de terceros a tampoco amplía la información que ya obre en su poder, pues aquel ya probablemente conocerá el propósito del tratamiento de los datos y la identidad de sus responsables (que, de no serles conocida, tampoco le va a aportar gran ayuda en la identificación del sesgo sufrido), aunque la norma añade finalmente que *la información básica podrá incluir en estos casos las categorías de datos objeto de tratamiento y las fuentes de su procedencia*. En este supuesto, el análisis de las categorías de datos permitiría acceder al origen del algoritmo decisor, pero de una manera superficial, si no se comparte el *código fuente*, y, aun compartiéndolo, puede no resultar tampoco suficiente si no se comparten los datos que alimentan el algoritmo[60].

2. Intervención humana significativa

La intervención humana conectada con las decisiones automatizadas cuenta con un único referente normativo en nuestro derecho positivo: el art. 22 RGPD (y su homólogo en el derecho interno). A tenor de dicho precepto y su interpretación en los considerandos previos, el derecho de explicabilidad queda referido a los *servicios de contratación en red en los que no medie intervención humana alguna*, lo que evidencia una necesidad urgente de ampliar esta estrecha esfera de atención legislativa a la adopción de decisiones laborales automatizadas. Pero, por otra parte, aun admitiendo su interpretación expansiva a otros ámbitos distintos de los servicios de contratación en red, considerando el uso que las empresas realizan de distintas herramientas con soporte de IA para adoptar decisiones (incluida la selección de personal y contratación, no en línea o en red, aunque esta sí sirva para desplegar el proceso de selección o incluso únicamente el de captación de la demanda de empleo, v.gr. cribado de currículos), el

habilitantes para el tratamiento de datos personales especialmente protegidos -es decir, a la identificación de los fines de interés público esencial y la apreciación de la proporcionalidad del tratamiento al fin perseguido, respetando en lo esencial el derecho a la protección de datos- como al establecimiento de "medidas adecuadas y específicas para proteger los intereses y derechos fundamentales del interesado" [art. 9.2 g) RGPD].

[60] Huergo Lora, A.: "Una aproximación a los algoritmos desde el Derecho administrativo", en Huergo Lora, A. (dir.) y Díaz González, G.M.: *La regulación de los algoritmos*. Aranzadi, Cizur Menor, 2020, pp. 54 y ss.; y Roig, A.: L*as garantías frente a las decisiones automatizadas: del Reglamento General de Protección de Datos a la gobernanza algorítmica*. Bosch, Barcelona, 2020.

concepto de intervención humana debe ser nuclear en su regulación y análisis jurídico. En particular, el concepto de intervención humana (o human-in-command), reivindicado también desde los planos ético y tecnológico, precisa de una acotación normativa que evite soslayar su exigibilidad en contextos laborales. Es decir, que permita excluir convalidaciones automáticas de recomendaciones automatizadas. En otras palabras, que implique que la intervención humana aludida sea algo más que procesar la orden de confirmación, que, por el contrario, tenga carácter substancial o significativo.

Ahora bien, caben dos planteamientos en este contexto. En primer lugar, el de la *intervención humana significativa*. Esta ha de poder explicar que el procesamiento automatizado previo a la decisión humana lo ha sido de aspectos accesorios o en los que no se adopte decisión final definitiva para ningún individuo. Utilizando un símil procedimental, se referiría a los actos que ponen fin a la vía de que se trate, v.g. de acceso a un puesto de trabajo (en lugar de al proceso o al procedimiento). En segundo lugar, el de la *acotación del concepto de decisión*. En este caso, se trataría de incluir en dicho ámbito de garantías no solo las decisiones que puedan considerarse "finales" o definitivas, v.gr., la decisión de contratación, sino todas aquellas que tengan relevancia para los sujetos implicados, como es la exclusión del proceso de contratación, lo que requeriría que la fase de cribado también fuera supervisada por humanos, quienes, además, asumieran la responsabilidad de los actos de notificación a los interesados, expresando de este modo la explicabilidad exigible, y facilitando la impugnación por parte de estos. En todo caso, no parece descabellado sostener la necesidad de que ambas vertientes de esta aproximación se incorporaran a los modelos de decisión automatizada en el ámbito del trabajo, porque se trata de decisiones que pueden tener afectación de derechos fundamentales como el derecho a la igualdad y a la no discriminación (lo que exigiría que una hipotética regulación específica de los derechos digitales laborales, en la línea de la Carta de Derechos Digitales aprobada por el Gobierno de España en julio de 2021, tuviera rango de ley orgánica).

3. Acceso a la motivación y derechos de propiedad intelectual

En el marco de la tutela antidiscriminatoria y, en particular, la justificación objetiva y razonable que puede salvar el indicio discriminatorio, en consecuencia, a la motivación de la decisión impugnada y la delimitación del alcance del derecho a una intervención humana (art. 22 RGPD), puede ser esencial determinar hasta dónde es posible acceder a ese instrumento intangible en el que se basó la decisión empresarial (secuenciación en la que consiste el algoritmo), como a los datos que lo alimentaron (librerías o información en forma de macrodatos). Sin embargo, la tendencia incipiente es centrar las reclamaciones contra decisiones automatizadas en el acceso a uno solo de tales elementos, al *código fuente*, con el propósito de verificar los defectos de diseño que puedan determinar el perjuicio que sostiene la reclamación (y conocer cómo se adoptó la decisión). No obstante, el abordaje jurídico de la cuestión no puede ser unívoco y monolítico, pues pueden concurrir dos situaciones diferenciadas: a) algoritmos de arquitectura más simple, no basados en aprendizaje automático, donde los datos de alimentación devienen secundarios, como es el caso planteado contra las aplicaciones administrativas para la solicitud de ayudas públicas (en el que el algoritmo debe solo determinar si los solicitantes cumplen o no con los criterios previamente introducidos para diseñar el algoritmo, v.g. el algoritmo BOSCO utilizado para la concesión del bono social eléctrico); o, b) los basados en aprendizaje automático o alimentación por datos, caso de los algoritmos predictivos usados en la selección de personal, donde los datos son precisamente la clave del aprendizaje del algoritmo para realizar su predicción o

selección (v.g. el algoritmo Send@ del Servicio Público de Empleo Estatal). Mientras en el primer caso seguramente conocer el código fuente o secuencia de programación permitirá adentrarse en el origen de la decisión (o su eventual manipulación sesgada), en el segundo caso, dicho acceso seguramente no aportará la base necesaria para plantear una oposición solvente al perjuicio causado por el sesgo).

Pues bien, las primeras reclamaciones en el plano laboral se han centrado únicamente en el *código fuente* (cfr. trabajadores de de Glovo[61] o de Uber, que reclamaban el acceso al "código fuente" del algoritmo), como forma de garantizar el derecho de transparencia (aunque dicho acceso no garantiza su inteligibilidad, pues precisa de conocimientos técnicos suficientes para su interpretación). De ahí los términos introducidos en el art. 64.4 d) ET por la Ley 12/2021, de 28 de septiembre, respecto de los derechos informativos de la representación legal del personal de la empresa. Sin embargo, si el *software* patentado o los algoritmos son propiedad intelectual de sus creadores (o adquirentes, si se cedieron sus derechos), es posible que la empleadora usuaria del mismo no ostente titularidad ni poder de disposición alguno (v.g. empresas adquirentes de una licencia de uso) que permita materializar el derecho de transparencia.

Por lo que respecta a los derechos de propiedad intelectual que puede oponer la empleadora, el Texto Refundido de la Ley de Propiedad intelectual (Real Decreto Legislativo 1/1996, de 12 de abril) lo identifica en su art. 96.1 con un "programa de ordenador" (lo será la operación o secuencia matemática en que consiste un algoritmo), susceptible de titularidad o propiedad por parte de personas físicas, dentro del contexto de los derechos de propiedad intelectual. Su amparo bajo derechos de propiedad intelectual se despliega sobre "las diferentes partes que integran una obra, (…) siempre que contengan determinados elementos que expresen la creación intelectual del autor (STJUE, Infopaq International, C-5/08, Rec. p. I-6569, apartado 39, y STJUE, Gran Sala, de 2 de mayo de 2012, asunto SAS Institute Inc contra World Programming Ltd., apartado 65). Y, asimismo, si el programa o algoritmo formara parte de una patente o un modelo de utilidad, gozará "de la protección que pudiera corresponderles por aplicación del régimen jurídico de la propiedad industrial" (art. 96.3). Dicha protección no se extiende sobre "las ideas y principios en los que se basan cualquiera de los elementos de un programa de ordenador incluidos los que sirven de fundamento a sus interfaces" (art. 96.4), de acuerdo con la Directiva 2009/24/CE del Parlamento Europeo y del Consejo, de 23 de abril de 2009, sobre la protección jurídica de programas de ordenador, que en su considerando undécimo se refiere explícitamente a los algoritmos: "de acuerdo con este principio de derechos de autor, en la medida en que la lógica, los algoritmos y los lenguajes de programación abarquen ideas y principios, estos últimos no están protegidos con arreglo a la presente Directiva", que deberán protegerse "mediante derechos de autor" en las legislaciones nacionales. Tal protección, conforme al considerando 63 del Reglamento UE de protección de datos personales, permite ocultar el código fuente en el contexto de una reclamación por cualquier interesado.

[61] En el primer caso, el Tribunal de Ámsterdam en sentencia C/13/692003/HA RK 20-302 de 11/3/2021 (y C/13/687315/HA RK 20-207 de la misma fecha) resuelve desfavorablemente, en tanto que se proporcionó a los trabajadores una explicación suficiente sobre dicho funcionamiento y esta no fue impugnada por ellos (sin perjuicio de la obligación de proporcionar acceso a los datos personales a los afectados). El segundo fue resuelto por sentencia del Tribunal ordinario de Bolonia de 27/11/2020, que estimó que los algoritmos utilizados por la empresa no eran inclusivos.

Ahora bien, el art. 96.2 establece que "el programa de ordenador será protegido únicamente si fuese original, en el sentido de ser una creación intelectual propia de su autor" (sustitúyase "programa" por "algoritmo"). A tenor del art. 97, solo son propietarios sus *creadores* o la persona jurídica a la que la ley reconozca tal titularidad, incluyendo a las empresas para la que presten sus servicios los trabajadores cuando su creación tenga lugar en el marco de su trabajo (en tal caso, el art. 97.4 atribuye al empresario la titularidad, en exclusiva, de los derechos de explotación del programa salvo pacto en contrario). Este efecto se infiere también, según las sentencias de la Audiencia Provincial de Soria núm. 136/2017, de 18 de octubre, y de la Audiencia Provincial de Valencia, núm. 164/2006, de 13 marzo, de la concertación de un arrendamiento de obra consistente en la creación de un programa informático a medida del cliente, que conlleva la obligación de entrega de contraseñas y del código fuente, por lo que, en caso de que el encargo se externalice con terceros, la empresa sigue siendo la propietaria del código fuente.

Por su parte, el art. 100 de la ley prevé algunas excepciones a los derechos de explotación, para salvar la necesaria autorización del titular (del algoritmo), pero entre ellos no se encuentra el simple acceso al código fuente, sino diversos modos de utilización o interacción, y, en todo caso, siempre que no se "perjudique(n) de forma injustificada los legítimos intereses del titular de los derechos" (art. 100.7). En cualquier caso, la mera observación, estudio o verificación del funcionamiento de un programa sin autorización previa del titular no constituye una infracción del derecho (Sentencia núm. 19/2019, de 18 enero, de la Audiencia Provincial de Madrid), conforme al art. 5.3 de la Directiva 2009/24/CE, pero, a tenor de la jurisprudencia de la UE, esta simple observación no es identificable con el acceso al código fuente.

En la misma línea, si la empresa hubiera obtenido una copia con licencia del algoritmo en cuestión, conforme a la Directiva 2009/24/CE, estaría asimismo autorizada a "observar, estudiar o verificar el funcionamiento de un programa de ordenador con el fin de determinar las ideas y los principios implícitos en cualquier elemento del programa", porque estas no están protegidas por los derechos de autor cubiertos por la directiva (STJUE Gran Sala, de 2 de mayo de 2012, asunto SAS Institute Inc contra World Programming Ltd, razonamiento 50). Igualmente el razonamiento 61 de la sentencia identifica esta situación con el mero uso del programa, sin acceso al código fuente, distinguiendo entre ello y estudiar, observar, verificar…, para concluir que "las palabras clave, la sintaxis, los comandos y combinaciones de comandos, las opciones, los valores por defecto y las iteraciones están compuestos por palabras, cifras o conceptos matemáticos que, considerados aisladamente, no constituyen, en cuanto tales, una creación intelectual del autor del programa de ordenador" (apdo. 66), aunque "sólo a través de la elección, la disposición y la combinación de tales palabras, cifras o conceptos matemáticos puede el autor expresar su espíritu creador de manera original y obtener un resultado, el manual de utilización del programa de ordenador, que constituye una creación intelectual (STJUE de 16 de julio de 2009, Infopaq International, C-5/08, Rec. p. I-6569, apdo. 39). En conclusión, aunque la resolución se refiera a otro núcleo de análisis (la copia de un programa informático o parte de él), el tribunal considera que esa combinación a la que llamamos algoritmo sí constituye una creación intelectual, que se documenta y escribe en lenguaje de código, pues "el objeto de la protección conferida por esa Directiva abarca el programa de ordenador en todas sus formas de expresión, que permiten reproducirlo en diferentes lenguajes informáticos, tales como el código fuente y el código objeto" (STJUE de 22 diciembre 2010, *asunto Bezpecnostní softwarová asociase*, apdo. 35).

Finalmente, la Ley 1/2019, de 20 de febrero, de secretos empresariales de secretos empresariales permite su acceso a la representación legal de los trabajadores (art. 2 c*)*, en correspondencia con la Directiva 2016/943, sobre secretos comerciales, art. 5), mientras que la Ley de Propiedad Intelectual admite su exposición en el contexto de una reclamación administrativa o judicial (art. 135).

C. Indicios y prueba de la discriminación algorítmica

1. Acreditación de los indicios de discriminación en caso de sesgos algorítmicos

Para que pueda evaluarse el carácter discriminatorio de una decisión, constituye presupuesto necesario la acreditación del indicio de la discriminación, a tenor de las reglas probatorias (art. 8 Directiva 2000/43/CE, art. 10 Directiva 2000/78/CE y art. 19 Directiva 2006/54/CE), aunque resulta difícil responder a la pregunta de si realmente decidir por mecanismos digitales interpuestos dificulta o no la prueba del indicio.

Como hipótesis de partida no cabe rechazar *a priori* que la interpretación de la apariencia de discriminación o indicio suficiente[62] *-prima facie-* no quede alterada como consecuencia del uso de un algoritmo, pues estos se presentan precisamente como herramientas para la objetividad y precisión en el asesoramiento de decisiones. En un escenario como el de un proceso de selección de trabajadores, el sesgo del algoritmo, de existir, puede acreditarse por mecanismos probatorios tradicionales (v.gr. comparación entre los individuos elegidos y los excluidos). Ahora bien, admitido el indicio, y considerando la gran precisión con la que puede operar el algoritmo de selección, a menos que el sesgo se encuentre en su propio diseño (código fuente), ¿cómo se efectúa el análisis comparativo en el universo objeto de procesamiento por el mismo frente al individuo excluido de la selección? En otras palabras, si en un proceso de selección con la concurrencia de cientos de candidatos son excluidos grupos de individuos distintos con características diversas protegidas (o incluso no protegidas), la comparación entre la persona reclamante y las no descartadas no permitirá constatar que la causa de exclusión sea únicamente la que esta posee, pues confluirá todo un universo de características inferidas que han podido ser igualmente rechazadas).

Si debe analizarse el binomio hombres-mujeres, y la sistemática exclusión de las mujeres, el análisis parece simple, pero si la base es el rechazo a características correspondientes a distintos grupos (discapacidad, etnia, religión, procedencia, entre otros) y no otros o no su combinación con otros, el análisis deja de aparentar simplicidad, en tanto la multitud de variables combinadas por el modelo automatizado permite considerar también que han primado otras características que asimismo deberán identificarse y que podrían pasar desapercibidas en la acreditación del indicio necesario. Seguidamente, resulta aún más complicado calificar como discriminatorias decisiones a cuando los criterios empleados por el algoritmo no correlacionan exactamente con

---

[62] Carrizosa, E.: "La concreción de los indicios de discriminación en la jurisprudencia comunitaria: STJUE 19 abril 2012", *Aranzadi Social: Revista Doctrinal*, vol. 5, núm. 7, 2012, pp. 59-65. Sobre la prueba de indicios, vid. Montoya Melgar, A.: "De sospechas e indicios en la discriminación antisindical (En torno a la STC 84/2002, de 22 de abril)", en AA.VV.: *Derecho vivo del trabajo y constitución: estudios en homenaje al profesor doctor Fernando Suárez González*. Tirant lo Blanch, Valencia, 2003, pp. 215-226; y Godínez Vargas, A.: "La carga de la prueba y la prueba de indicios como medidas compensatorias de equilibrio procesal en juicios de discriminación", *Derecho laboral: Revista de doctrina, jurisprudencia e informaciones sociales*, núm. 264, 2016, pp. 671-688.

características atribuibles a una categoría social protegida, aunque pueden relacionarse igualmente con tal característica. Por ejemplo, tomando como referencia el caso citado por Xenidis[63], si el fundamento de la decisión del algoritmo se basa en variables como la distancia del lugar de trabajo, esta no determina *per se* la existencia de un criterio discriminatorio bajo el derecho positivo, pero si tal factor se anuda a una procedencia concreta, y esta a una característica protegida, la inferencia realizada por el algoritmo deriva en una decisión discriminatoria (v.g. los individuos que residen en la zona rechazada son mayoritariamente población inmigrante). Si, a su vez, la característica con la que se realiza la asociación no se encuentra entre las protegidas por el derecho antidiscriminatorio, se podría producir una situación de discriminación interseccional producto de la confluencia de varios factores que solo un análisis más exhaustivo podría visibilizar y que únicamente un marco de protección más amplio que el proporcionado por las directivas o el art. 14 CE podría calificarse de discriminatorio. Ahora bien, lo que también es importante tener en cuenta es la posible ruptura del nexo de conexión necesario para establecer la preceptiva relación entre la decisión y la discriminación pretendida, precisamente por la distancia entre los datos de referencia y la consecuencia analizada. En definitiva, lo más relevante es que la propia dinámica de funcionamiento de los algoritmos introduce una dosis mucho mayor de dificultad en la detección de sesgos discriminatorios, que obliga a perfeccionar la granularidad del análisis y reconsiderar la estrategia jurídica frente a la tutela de la discriminación producto de mecanismos de decisión automatizada.

2. En caso de discriminación múltiple y/o interseccional

La discriminación múltiple e interseccional encuentran una directa equivalencia entre los métodos de inferencia de datos y el resultado discriminatorio, lo que significa que la admisión legal de la autonomía de ambas figuras permitiría dar justa cobertura a bolsas de discriminación que, por acción de los estudiados mecanismos automatizados, no solo pueden estar siendo objeto de un incremento exponencial en buena parte invisible, sino que, además, permanecen al margen de la adecuada tutela jurídica.

Descendiendo al plano probatorio, resulta de interés analizar si la equivalencia entre la interseccionalidad y el sesgo resultante de la inferencia entre campos de datos es susceptible de acreditación probatoria. En la medida en que el aprendizaje automático no permite conocer la correlación entre datos y cuáles de ellos han determinado el resultado, es decir, no es posible saber cuáles de los rasgos analizados han sido determinantes para el resultado ofrecido por el mecanismo automatizado, y si técnicamente no es factible asistir esta explicación a la hipotética víctima de la decisión discriminatoria, la consecuencia es que probablemente estemos ante un caso de discriminación múltiple, interseccional, pero no sea posible constatarlo, salvo que la empleadora proporcione también a los datos de contraste (no los datos de alimentación). En definitiva, se trata del esquema tradicional de aportación de indicios, facilitado por el uso de otro sistema automatizado capaz de hallar la correlación entre un conjunto de individuos y el individuo que alega la discriminación, que bien pudieran también ser asistidos por mecanismos automatizados capaces de detectar la incidencia estadística simple o interseccional de sesgos discriminatorios.

[63] Xenidis, R.: "Tuning EU equality law to algorithmic discrimination…", cit., p. 4.

D. A modo de conclusión

Las deficiencias observadas en el vigente derecho antidiscriminatorio, especialmente ligadas a la falta de admisión en el derecho positivo de la doctrina de la discriminación múltiple y por intersección, dificultan la tutela efectiva del derecho a la no discriminación cuando esta se funda en decisiones automatizadas, por cuanto estas usan un sistema de prejuicios basados en la realidad social que absorben, pero que gestionan y procesan conforme a parámetros propios. Esa diferencia de planos de análisis y lenguajes motiva la desatención a un amplio núcleo de situaciones potencialmente discriminatorias, lo que justifica la necesidad real y no meramente proyectiva e hipotética de una mejora sustancial de los conceptos jurídicos que definen los distintos parámetros de la discriminación laboral.

En esta línea, tanto la positivización explícita de los distintos derechos que acompañan a la tutela frente al uso de datos masivos (*explicabilidad* y acceso al renacimiento subyacente, intervención humana *significativa* o relevante, y también las garantías frente al "blindaje" de la propiedad intelectual que pueda emplearse para encubrir decisiones poco claras), como su adaptación autónoma a las decisiones automatizadas, así como la admisión del haz conceptual de la discriminación múltiple y por intersección en nuestro derecho permitirían una clarificación del entorno jurídico en el que se están aplicando decisiones automatizadas con impacto directo sobre derechos fundamentales y una mejor tutela de los trabajadores frente a la discriminación en el trabajo basada en tal automatización. Lo cierto es que ya existe una proposición de ley que apunta en esa línea (Proposición de Ley Integral de Igualdad de 2021), aunque sus referencias a la automatización de decisiones tienen una oportunidad única para incluir los matices apuntados en favor de la seguridad jurídica y, sobre todo, de la evitación de núcleos de desprotección en el ámbito laboral regido por inteligencia artificial.

Bibliografía

Barocas, S. y Selbst, A. D.: "Big data's disparate impact". *California Law Review*, núm. 104, 2016, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899.

Barrère Unzueta, M. A.: "La interseccionalidad como desafío al mainstreaming de género en las políticas públicas", *Revista Vasca de Administraciones Públicas*, núm. 87-88, 2010.

Barzilay, A. y Ben-David, A.: "Platform Inequality: Gender in the Gig-Economy, "*Seton Hall Law Review*, vol. 47, núm. 2, 2017.

Burrell, J.: "How the machine 'thinks': understanding opacity in machine learning algorithms". Vol. 3, núm. 1, 2016, https://doi.org/10.1177/2053951715622512.

Carrizosa, E.: "La concreción de los indicios de discriminación en la jurisprudencia comunitaria: STJUE 19 abril 2012", *Aranzadi Social*, vol. 5, núm. 7, 2012, pp. 59-65.

Castán Tobeñas, J.: *La idea de equidad y su relación con otras ideas morales y jurídicas afines*, Madrid, Reus, 1950.

Castellanos Claramunt, J., y Montero Caro, M. D.: "Perspectiva constitucional de las garantías de aplicación de la inteligencia artificial: la ineludible protección de los derechos fundamentales". *Ius et Scientia*, vol. 6, núm. 2, 2020, https://dx.doi.org/10.12795/IETSCIENTIA, pp. 72-82.

Chouldechova, A.: "Fair prediction with disparate impact: a study of bias in recidivism prediction instruments", 2016, pp. 1-17, en https://arxiv.org/abs/1610.07524.

Crenshaw, K.: "Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics". *The University of*

Chicago Legal Forum: 1989. HeinOnline - 1989 U. Chi. Legal F. 139 1989, https://philpapers.org/archive/CREDTI.pdf?ncid=txtlnkusaolp00000603.

De la Sierra Morón, S.: "Control judicial de los algoritmos: robots, administración y estado de derecho". Lefebvre, Derecholocal.es, 21/5/2021, https://derecholocal.es/opinion/control-judicial-de-los-algoritmos-robots-administracion-y-estado-de-derecho.

Ebers, M.: "Ethical and legal challenges", en Ebers, M., y Navas, S., dirs.: *Algorithms and law*. Cambridge University Press, 2020, pp. 37-99.

Gerards, J. y Xenidis, R.: "Algorithmic discrimination in Europe: Challenges and Opportunities for EU equality law", *European Futures*, 3/12/2020, https://www.europeanfutures.ed.ac.uk/algorithmic-discrimination-in-europe-challenges-and-opportunities-for-eu-equality-law/.

Godínez Vargas, A.: "La carga de la prueba y la prueba de indicios como medidas compensatorias de equilibrio procesal en juicios de discriminación", *Derecho laboral: Revista de doctrina, jurisprudencia e informaciones sociales*, núm. 264, 2016, pp. 671-688.

Gonzalo Quiroga, M.: "Discriminación racial y control de identificación policial: valoración de la raza como indicio de extranjería y de nacionalidad", *La Ley: Revista jurídica española de doctrina, jurisprudencia y bibliografía*, núm. 3, 2001, pp. 2158-2164.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E. y Nelson, C.: "Clinical versus mechanical prediction: a meta-analysis". *Psychological assessment*, vol. 12, núm. 1, 2000.

Gunning, D.: "Explainable Artificial Intelligence (XAI)", 2017, https://www.cc.gatech.edu/~alanwags/DLAI2016/(Gunning)%20IJCAI-16%20DLAI%20WS.pdf.

Hajian, S.: *Simultaneous discrimination prevention and privacy protection in data publishing and mining*, tesis doctoral en filosofía de ciencia computacional, Universidad Rovira i Virgili, 2013, https://www.tdx.cat/bitstream/handle/10803/119651/thesis.pdf?sequence=1.

Hajian, S., Bonchi, F., y Castillo, C.: "Algorithmic Bias: From Discrimination Discovery to Fairness-aware Data Mining", DOI: 10.1145/2939672.2945386, Conference: the 22nd ACM SIGKDD International Conference, KDD, 2016, disponible en https://www.researchgate.net/publication/305997939_Algorithmic_Bias_From_Discrimination_Discovery_to_Fairness-aware_Data_Mining.

Hajian, S. y Ferrer, J. D.: "A Methodology for Direct and Indirect Discrimination Prevention in Data Mining", IEEE Transactions on Knowledge and Data Engineering, 2013, DOI: 10.1109/TKDE.2012.72.

Hildebrandt, M.: "Algorithmic regulation and the rule of law". *Philosophical Transactions of the Royal Society A*, vol. 376, núm. 2128, 2018. DOI:http://dx.doi.org/10.1098/rsta.2017.0355.

Hildebrandt, M.: "The issue of bias. The framing powers of ML". *Computer Science*, DOI:10.2139/ssrn.3497597. En M. Pelillo, T. Scantamburlo (eds.): *Machine We Trust. Perspectives on Dependable AI*, MIT Press 2021, http://dx.doi.org/10.2139/ssrn.3497597. Preprint.

Ho, D. E., y Xiang, A.: "Affirmative Algorithms: The Legal Grounds for Fairness as Awareness". *The University of Chicago Law Review Online* (30/10/2020), https://lawreviewblog.uchicago.edu/2020/10/30/aa-ho-xiang/.

Huergo Lora, A.: "Una aproximación a los algoritmos desde el Derecho administrativo", en Huergo Lora, A. (dir.) y Díaz González, G.M.: *La regulación de los algoritmos*. Aranzadi, Cizur Menor, 2020.

Kayser-Bril, N.: "Spain: Legal fight over an algorithm's code", *Algorithwatch*, 12/8/2019, en https://algorithmwatch.org/en/story/spain-legal-fight-over-an-algorithms-code/.

López Rodríguez, A.M.: *Derecho comparado y digitalización*. Tecnos, Madrid, 2021.

Makkonen, T.: *Multiple, Compound and Iintersectional Discrimination: bringing the experiences of the most marginalized to the fore. Institute For Human Rights*, Abo Akademi University. 2002.

Mantelero, A.: *Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data*, Consejo de Europa, T-PD(2017)01, en https://rm.coe.int/16806ebe7a.

Mayson, S.G.: "Bias In, Bias Out". *The Yale Law Journal*, vol. 128, núm. 8, 2019, https://www.yalelawjournal.org/article/bias-in-bias-out#:~:text=abstract.,to%20have%20disparate%20racial%20impacts.

Miné, M.: "Los conceptos de discriminación directa e indirecta", Conferencia "Lucha contra la discriminación: Las nuevas directivas de 2000 sobre la igualdad de trato", 31/3-1/4/2003, Trier, http://www.era-comm.eu/oldoku/Adiskri/02_Key_concepts/2003_Mine_ES.pdf

Montoya Melgar, A.: "De sospechas e indicios en la discriminación antisindical (En torno a la STC 84/2002, de 22 de abril)", en AA.VV.: *Derecho vivo del trabajo y constitución: estudios en homenaje al profesor doctor Fernando Suárez González*. Tirant lo Blanch, Valencia, 2003, pp. 215-226.

O'Neil, C.: *Armas de destrucción matemática*. Capitán Swing, Madrid, 2017.

Pasquale, F.: "A Rule of Persons, Not Machines: The Limits of Legal Automation", *The George Washington Law Review*, vol. 87, núm. 1, 2019, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3135549.

Pasquale, F.: *New laws of robotics: defending human expertise in the age of AI*. The Belknap Press, 2020.

Rey Martínez, F.: "La discriminación múltiple, una realidad antigua, un concepto nuevo". *Revista española de derecho constitucional*, año 28, núm. 84, 2008, pp. 251-283.

Rivas Vallejo, P.: *La aplicación de la inteligencia artificial al trabajo y su impacto discirminatorio*. Thomson Reuters Aranzadi, Cizur Menor, 2020.

Roig, A.: *Las garantías frente a las decisiones automatizadas: del Reglamento General de Protección de Datos a la gobernanza algorítmica*. Bosch, Barcelona. 2020.

Rosenblat, A.: *Uberland Cómo los algoritmos están reescribiendo las reglas de trabajo.* University of California Press, 2018.

Ruiz-Gallardón, I.: "La equidad: una justicia más justa". *Foro, Nueva época*, vol. 20, núm. 2, 2017, pp. 173-191, http://dx.doi.org/10.5209/FORO.59013.

Schiek, D. y Lawson, A. (dirs.): *European Union Non-Discrimination Law and Intersectionality: Investigating the triangle of racial, gender and disability discrimination*, Londres-Nueva York: Routledge, 2016.

Serra Cristóbal, R. (coord.): *La discriminación múltiple en los ordenamientos jurídicos español y europeo*, Valencia: Tirant lo Blanch, 2013.

Serra Cristóbal, R.: "El reconocimiento de la discriminación múltiple por los tribunales". *Teoría y derecho*, núm. 27, 2020, pp. 140-161. DOI: https://doi.org/10.36151/td.2020.008.

Smith-Strother, L.: "The role of social advocacy in diversity & inclusion recruiting", Glassdoor Summit 2016, https://youtu.be/IdsqQMV4V_0).

Stoica, A.-A., Riederer, C. y Chaintreau, A.: "Algorithmic glass ceiling in social networks: the effects of social recommendations on network diversity". *Proceedings of the Web Conference* 2018, Lyon. ACM, Nueva York, pp. 923–932, https://doi.org/10.1145/3178876.3186140.

Tomei, M.: "Análisis de los conceptos de discriminación y de igualdad en el trabajo". *Revista Internacional del Trabajo*, vol. 122, núm. 4, 2003.

Vantin, S.: “Inteligencia artificial y derecho antidiscriminatorio”, en Llano Alonso, F. y Garrido Martín, J. (eds.): *Inteligencia artificial y derecho. El jurista ante los retos de la era digital*. Thomson Reuters Aranzadi, Cizur Menor, 2021.

Xenidis, R.: “Tuning EU equality law to algorithmic discrimination: three pathways to resilience”, *Maastricht Journal of European and Comparative Law* 2020, vol. 27, núm. 6, 4/1/2021, en https://doi.org/10.1177/1023263X20982173.

Xenidis, R. y Senden, L.: “EU non-discrimination law in the era of artificial intelligence: mapping the challenges of algorithmic discrimination”. U. Bernitz et al (eds): *General principles of EU law and the eu digital order*. Kluwer Law Int., 2020, pp. 151-182.

Wachter, S.: “Affinity profiling and discrimination by association in online behavioural advertising”. *Berkeley Technology Law Journal*, núm. 35, 2020, https://btlj.org/data/articles2020/35_2/01-Wachter_WEB_03-25-21.pdf.

Zuiderveen Borgesius, F.: *Discrimination, artificial intelligence, and algorithmic decision-making*. Council of Europe, Directorate General of Democracy, 2018.

NOTA: todos los enlaces digitales fueron verificados en el mes de diciembre de 2021.