DISCRIMINACIÓN ALGORÍTMICA: ¿DIRECTA O INDIRECTA? UN ESTUDIO SOBRE LA INADECUACIÓN DE ESTA CONCEPCIÓN BIDIMENSIONAL

Algorithmic discrimination: Direct or indirect? A study on the inadequacy of this two-dimensional understanding

ANNA CAPELLÀ RICART
Universitat Autònoma de Barcelona
anna.capella@uab.cat

Cómo citar/Citation
Capellà Ricart, A. (2024).
Discriminación algorítmica: ¿directa o indirecta? Un estudio sobre la inadecuación de esta concepción bidimensional
IgualdadES, 11, 67-96
doi: https://doi.org/10.18042/cepc/IgdES.11.03

(Recepción: 12/04/2024; aceptación tras revisión: 07/10/2024; publicación: 13/12/2024)

Resumen

El derecho antidiscriminatorio europeo está estructurado de acuerdo con la apreciación de la discriminación según si esta es directa o indirecta. Existe un debate doctrinal sobre en qué tipología se puede integrar mejor la discriminación causada por sistemas algorítmicos, que teniendo en cuenta las particularidades que presentan estos, como la ininteligibilidad, la consideración de multiplicidad de atributos protegidos o la absorción de sesgos implícitos, se inclina por la discriminación indirecta. Sin embargo, planteamos si no sería más interesante cambiar de perspectiva dadas las dificultades para dilucidar frente a qué tipo de discriminación estamos y focalizar la apreciación de discriminación sobre la base de la concreción de ciertos estándares de precisión según los colectivos sociales (incluyendo categorías protegidas combinadas) que se pueden ver afectados por el sistema, siendo los resultados relativos a la precisión el punto de partida para apreciar si existe o no discriminación.

Palabras clave

Discriminación directa; discriminación indirecta; inteligencia artificial; precisión; Unión Europea.

Abstract

European anti-discrimination law is structured based on the assessment of discrimination according to whether it is direct or indirect. There is a doctrinal debate about what typology can address better the discrimination caused by algorithmic systems that, considering this kind of systems' particularities, such as unintelligibility, consideration of multiple protected attributes, or absorption of implicit biases, leans for indirect discrimination. However, we consider whether it would not be more interesting to change perspective given the difficulties in elucidating what type of discrimination we are facing and focus the appreciation of discrimination on assessing if the system achieves or not certain precision standards according to particular groups of population (including groups created by combined protected attributes) that may be affected by the system, and using precision results as a starting point to appreciate whether discrimination has occurred.

Keywords

Direct discrimination; indirect discrimination; artificial intelligence; precision; European Union.

SUMARIO

I. INTRODUCCIÓN

Las decisiones tomadas por sistemas algorítmicos pueden generar resultados discriminatorios (véase, por ejemplo, Allhutter *et al.*, 2020). En la tradición europea si se trata una persona de manera menos favorable de lo que ha sido o de lo que vaya a ser tratada otra en una situación comparable a causa de una categoría protegida[1] por la legislación (raza o etnia, sexo, orientación sexual, religión o creencias, edad o discapacidad), se considera que se ha ocasionado discriminación directa[2]. En cambio, si un criterio o práctica aparentemente neutro discrimina a personas dotadas de un atributo protegido

[1] En este trabajo nos referiremos a *categoría protegida* o *atributo protegido* con el objetivo de englobar aquellas características personales que la legislación decide proteger especialmente (por ejemplo, etnia u origen étnico, sexo, edad, orientación sexual, religión o convicciones o discapacidad).

[2] Véase, por ejemplo: TEDH, sentencia de 2 de diciembre de 2014, 61960/08, *Emel Boyraz v. Turquia*; TJCE, sentencia de 10 de julio de 2008, *Firma Feryn NV*, C-54/07, EU:C:2008:397.

se considera discriminación indirecta[3]. A pesar de la aparición de otros tipos de discriminación (por asociación, por error, interseccional, estructural, sistémica), en el ámbito jurídico la conceptualización bidimensional entre discriminación directa e indirecta sigue siendo un elemento central para dilucidar casos de discriminación, ya que el hecho que la casuística recaiga en un tipo o en otro tiene implicaciones en relación con las justificaciones que se pueden plantear: en la discriminación directa son *numerus clausus* y, en cambio, en la discriminación indirecta no, y la determinación de si existe una justificación depende del análisis de cada caso (Tobler, 2022: 78).

En este articulo ponemos en duda que la distinción entre discriminación directa o indirecta sea un buen método para abordar jurídicamente la discriminación que se produce en base a resultados de sistemas algorítmicos. Justamente la crítica por las desigualdades que pueden generar los sistemas de inteligencia artificial (IA) aplicados a contextos sociales ha contribuido a evidenciar que la discriminación no es tanto un hecho puntual como un conjunto de dinámicas sociales y estructuras de poder que en su desarrollo social implican un trato menos favorable hacia ciertas personas que comparten ciertas características.

Leese planteó que las regulaciones antidiscriminación solo desplegaban la plenitud de su poder regulador cuando se aplicaban a modos tradicionales de discriminación estáticos, y no a aquellos provocados por el análisis masivo de datos, dinámico e implícito, que escapa constantemente a la regulación, poniendo en duda si se podía abordar mediante la perspectiva de la discriminación directa e indirecta establecida hasta ahora (Leese, 2014: 500).

Desde la perspectiva antisubordinatoria, que considera que el derecho antidiscriminatorio tiene como objetivo luchar contra la subordinación de ciertos grupos sociales frente a otros (en contraposición a la perspectiva anticlasificatoria, que tiene como finalidad luchar frente al trato desigual arbitrario y no razonable), se considera que la clasificación directa/indirecta se integró en el derecho europeo y cristalizó en un esquema binario que obstaculiza el avance del derecho antidiscriminatorio al eclipsar para el derecho las relaciones de poder intergrupal (Barrère, 2008: 64). Tal y como lo planteó Foucault (1979: 171), no conviene partir de un hecho primero y masivo de dominación (una estructura binaria compuesta por «dominantes» y «dominados»), sino que es necesario considerar la producción multiforme de relaciones de poder que son parcialmente integrables a estrategias de conjunto.

[3] Véase, por ejemplo: TEDH, sentencia de 2 de febrero de 2016, 7186/09, *Di Trizio v. Suiza*; TJCE; sentencia de 13 de mayo de 1986, *Bilka-Kaufhaus*, C-170/84, EU:C:1986:204.

Además, se ha alertado de que la concepción binaria de discriminación directa e indirecta dificulta la articulación de respuestas frente a la discriminación algorítmica (Añón Roig, 2022: 36 y 37).

Es relevante plantear que el TJUE, cuando analiza casos de discriminación, normalmente delega al tribunal nacional dilucidar si se trata de un caso de discriminación directa o indirecta[4]. Siguiendo a Tobler (2022: 88), la jurisprudencia del TJUE actualmente distingue la existencia de discriminación directa o indirecta de la siguiente manera: se tratará de discriminación directa cuando, o bien se ha hecho un uso expreso de una categoría protegida por la legislación (ver, por ejemplo, la sentencia *Mangold*[5]), o bien se puede probar que un criterio se ha escogido por razones relacionadas con una categoría protegida por la legislación, a pesar de ser formalmente neutro (véase la argumentación que hace el tribunal en el asunto *CHEZ*), o bien el criterio usado es inseparable o está inextricablemente vinculado a una categoría protegida por la legislación (véase, por ejemplo, el caso *Szpital Kliniczny*). Al parecer, cuando se trata de un criterio inseparable de la categoría protegida, el TJUE se focaliza en el efecto excluyente del criterio empleado: o bien todas las personas que se ven perjudicadas por el criterio forman parte del mismo grupo protegido (véase, por ejemplo, el asunto *Nikoloudi*[6]), o bien el grupo protegido incluye a personas que están completamente excluidas, en el sentido de que nunca podrán cumplir el criterio en cuestión (ver, por ejemplo, la sentencia *Hay*[7]). Por el contrario, se tratará de discriminación indirecta cuando se utilice cualquier otro criterio formalmente neutro, donde tanto el grupo aventajado como el grupo desaventajado sean heterogéneos (ver, por ejemplo, el asunto *Bilka-Kaufhaus*).

En este artículo estudiamos el abordaje de la discriminación algorítmica. En este sentido estudiamos el abordaje de la discriminación algorítmica, en primer lugar, a través de la doctrina de la discriminación directa (1); en segundo lugar, a través de la discriminación indirecta (2), analizando seguidamente la inadecuación, desde de nuestra perspectiva, de esa concepción bidimensional (3).

---

4 Véase, por ejemplo: TJUE, sentencia de 21 de julio de 2011, *Patrick Kelly*, C-104/10, EU:C:2011:506; TJUE, sentencia de 19 de abril de 2012, *Galina Meister*, C-415/10, EU:C:2012:217; TJUE, sentencia de 16 de julio de 2015, *CHEZ Razpredelenie Bulgaria*, C-83/14, EU:C:2015:480; TJUE, sentencia de 26 de enero de 2021, VL contra *Szpital Kliniczny*, C-16/19, EU:C:2021:64.

5 TJCE, sentencia de 22 de noviembre de 2005, *Weiner Mangold*, C-144/04, EU:C:2005:709.

6 TJCE, sentencia de 10 de marzo de 2005, *Vasiliki Nikoloudi*, C-196/02, EU:C:2005:141.

7 TJUE, sentencia de 12 de diciembre de 2013, *Frédéric Hay*, C-267/12, EU:C:2013:823.

II. DISCRIMINACIÓN ALGORÍTMICA: UNA FORMA DIRECTA DE DISCRIMINACIÓN

Se define la discriminación directa como aquella situación en la que una persona es tratada de forma menos favorable de lo que ha sido o vaya a ser tratada otra persona en una situación comparable a causa de una categoría protegida (Rey Martínez, 2019: 56).

1. ELEMENTOS QUE CONFORMAN LA DISCRIMINACIÓN DIRECTA

Para determinar que existe discriminación directa, el TEDH requiere una diferencia de tratamiento de personas en situaciones análogas o significativamente similares, basada en una característica identificable. Para justificar la discriminación directa, en el caso del TEDH solo se requiere que se persiga una finalidad legítima y que los medios para conseguirla sean adecuados y necesarios. En el caso del derecho de la UE, no existe una justificación objetiva general como en el TEDH, sino que el tratamiento más desfavorable debe justificarse en base a las excepciones relativas a requisitos profesionales esenciales (Agencia Europea de Derechos Fundamentales y Consejo de Europa, 2019: 105, 106). Estas excepciones permiten a aquellas personas que contraten diferenciar a las personas según los atributos que poseen y que están especialmente protegidos por la legislación (por ejemplo, establecer condiciones mínimas en lo que se refiere a la edad para el acceso a un empleo[8]).

El concepto de discriminación directa tiene ciertas limitaciones que Fredman (2011: 166) concreta en ser «relativo» (requiere normalmente un comparador apropiado) y «simétrico» (también puede ser invocado por miembros de la mayoría social en caso de recibir un trato diferente y peor que los de una minoría social, solo por pertenecer a la mayoría)[9].

Se trata de una forma de discriminación que puede darse cuando se ven implicados sistemas algorítmicos; por ejemplo, cuando estos deciden que las mujeres no son adecuadas para un puesto de trabajo, ya que el algoritmo ha sido entrenado con bases de datos donde no existen prácticamente currículos de mujeres. El algoritmo, pues, otorgará al hecho de ser mujer una valoración menos positiva en comparación con el hecho de ser hombre, debido a su infrarrepresentación en la base de datos, y llegará a la conclusión de que es

8 Art. 6.1.b de la Directiva 2000/78/CE del Consejo, de 27 de noviembre de 2000, relativa al establecimiento de un marco general para la igualdad de trato en el empleo y la ocupación.

9 La autora, cuando habla de simetría, se está refiriendo a la discriminación inversa.

preferible contratar a hombres. En consecuencia, el valor negativo (o menos positivo) asociado a la categoría de datos podrá determinar de forma directa el resultado (Soriano Arnanz, 2021a: 15).

En la discriminación directa, el sistema algorítmico trata una categoría protegida o la pertenencia a un grupo desaventajado como factor de entrada negativo para la decisión que adoptar. También puede darse cuando el propio algoritmo infiere la pertenencia a un grupo desaventajado a través de otros datos a los que atribuye un valor negativo, ya sea que las inferencias las articulen las personas que creen el programa, ya sea que el algoritmo las desarrolle mediante el aprendizaje automático discriminación de las mujeres debido a una base de datos integrada mayoritariamente por hombres (Soriano Arnanz, 2022: 150). Soriano Arnanz expone que es relevante si el dato es indicativo, pero no determinante de la pertenencia al grupo, puesto que en este caso sería dudoso si obtendría la consideración de discriminación directa (ya que el dato sería un factor equivalente al atributo protegido) o discriminación indirecta (medida discriminatoria que solo puede probarse a través de los resultados desventajosos por un grupo de población). Por ejemplo, haber ido a una escuela en la que la mayoría del alumnado pertenece a una minoría étnica implica una mayor probabilidad de formar parte de esta minoría étnica, pero no es una variable totalmente predictiva de la pertenencia a este colectivo si el alumnado no se compone exclusivamente de miembros de esta etnia. Otra cosa sería la inferencia de la pertenencia al grupo protegido a través de datos que sean inseparables de la pertenencia: si resulta que en la escuela únicamente acceden miembros de un determinado grupo étnico o racial, el dato será inseparable de la pertenencia en el grupo (*ibid.*: 150 y ss.).

La discriminación directa, además, puede darse en dos tipologías. O bien la pertenencia de una persona al grupo protegido es el único factor que se tiene en cuenta en la decisión y genera de forma automática un trato discriminatorio, o bien la pertenencia de una persona a un grupo protegido no es el único factor que se tiene en cuenta en la decisión y otros datos hacen que el resultado final no sea discriminatorio. En este segundo caso, existe una combinación de factores que hacen que el resultado discriminatorio relativo a la categoría protegida se vea compensado por otros factores. Por ejemplo, en el caso del sistema algorítmico basado en datos estadísticos del mercado laboral de la Agencia Estatal de Trabajo de Austria (AMS) ser mujer resta puntos, pero ser austríaca o nacional de la UE suma puntos. Cabe apuntar que en el caso de los sistemas algorítmicos donde básicamente existe análisis masivo de datos y correlación de datos es muy improbable que la decisión se base únicamente en una variable, así como será difícil determinar qué peso se da a las categorías protegidas y, si como resultado del peso que se les da, existe una afectación en

el resultado final —donde ha habido discriminación, pero el resultado final puede acabar siendo no discriminatorio— (*ibid.*: 152).

Es relevante indicar que en el Reglamento europeo de la IA recientemente aprobado se exigen ciertos requisitos a los sistemas de IA de riesgo alto (art. 9 a 15) para promover que sean más transparentes y comprensibles. Además, en el art. 86 del Reglamento se establece que cualquier persona en relación con la cual se haya tomado una decisión en base al resultado de un sistema de IA de riego alto podrá obtener una explicación clara y significativa del rol del sistema de IA en el proceso de decisión y de los elementos principales que han determinado la decisión. Aún queda por ver, sin embargo, si estos requisitos y derechos enunciados van a tener una incidencia real en la consecución de unos sistemas de IA más transparentes y menos discriminatorios.

En resumen, el problema principal de la vertiente directa de la discriminación algorítmica será probar que la tipología de discriminación ha sido directa. Como establece Benedi Lahuerta (2016: 807, 808) en relación con el asunto CHEZ, planteado ante el TJUE, en los casos de discriminación algorítmica será preferible demostrar la existencia de discriminación directa (a efectos de la restricción en la justificación posterior), pero tendrá que haber suficientes evidencias para demostrar que la discriminación se produjo debido a la característica protegida.

2. APUNTES DE LA JURISPRUDENCIA DE TJUE EN RELACIÓN CON LA DISCRIMINACIÓN DIRECTA

El TJUE ha sentenciado algunos casos de discriminación directa con corrientes interpretativas que pueden ser beneficiosas para abordar la discriminación algorítmica: cabe mencionar el caso *Dekker*[10] y el caso *Tadao Maruko*[11].

En primer lugar, en el caso *Dekker* se inició una modificación relativa a la exigencia de comparación en los litigios de discriminación por razón de embarazo, puesto que al no existir una situación comparable en el caso de los candidatos de sexo masculino (nunca podrían recibir una negativa de contratación basada en este motivo) se establecía el embarazo como factor indisociable al sexo femenino. Esta línea jurisprudencial parece favorable para abordar la discriminación algorítmica a raíz de factores equivalentes al atributo protegido.

[10] TJCE, sentencia de 8 de noviembre de 1990, *Elisabeth Johanna Pacifica Dekker*, C-177/88, EU:C:1990:383.

[11] TJCE (Gran Sala), sentencia de 1 de abril de 2008, *Tadao Maruko*, C-267/06, EU:C:2008:179.

Sin embargo, cabe apuntar que se puede aplicar bajo la condición de que exista una transparencia mínima en el sistema algorítmico que permita determinar basándose en qué factores se ha tomado la decisión, para poder así establecer si estos son o no son indisociables al atributo protegido. No está claro cuándo el vínculo entre los indicadores estadísticos (en inglés *proxys*) y un atributo protegido será considerado suficientemente consistente por parte del TJUE para considerar el caso como discriminación directa (Xenidis, 2020: 746). Tal y como en el caso *Dekker*, el TJUE trató la discriminación por embarazo como discriminación directa al establecer un vínculo indisociable entre el sexo femenino y el embarazo. En el caso *Jyske Finans*[12] determinó que el país de nacimiento no servía por sí solo para establecer una presunción general de pertenencia a un determinado grupo étnico, es decir, no consideró la existencia de un vínculo directo o indisociable entre el país de nacimiento y el origen étnico. En este asunto, el TJUE pidió más de un indicador del motivo protegido, considerando que el país de nacimiento solo sería uno de los factores específicos que permitirían concluir que una persona pertenece a un grupo étnico, sin ser el único (págs. 17 y 18).

En segundo lugar, en el caso *Tadao Maruko*, el TJUE debía responder si una disposición estatutaria de un organismo alemán con personalidad jurídica pública que gestionaba seguros era contraria al art. 2.2.a de la Directiva 2000/78/CE, al disponer que un miembro de una pareja inscrita, al morir el otro miembro de la pareja, no tenía derecho a percibir una pensión de supervivencia tal y como correspondería a un cónyuge, a pesar de haber mantenido una unión similar al matrimonio.

El abogado general, en sus conclusiones, afirmó que denegar —basándose en la disposición estatutaria— la pensión porque no había matrimonio (reservado a personas heterosexuales), cuando se había formalizado una unión, con efectos sustancialmente idénticos, entre personas del mismo sexo, suponía una discriminación indirecta por razón de orientación sexual (pár. 102). Según el abogado general, no se trataba de una discriminación directa, puesto que el rechazo no se fundaba en la orientación sexual de la persona interesada, sino que se fundaba en una disposición aparentemente neutra —distinción entre parejas inscritas y parejas casadas—, que ocasionaba una desventaja a personas homosexuales (pár. 102)[13].

---

[12] TJUE (Sala Primera), sentencia de 6 de abril de 2017, *Jyske Finans*, C-668/15, EU:C:2017:278.

[13] Conclusiones del abogado general Dámaso Ruiz-Jarabo Colomer, presentadas el 6 de septiembre de 2007, en el asunto *Tadao Maruko*, C-267/06, EU:C:2007:486, pár. 102.

Sin embargo, el TJUE en su sentencia, abrió la posibilidad de que el tribunal nacional valorara si se trataba de discriminación directa al establecer que, si solo se podía ser pareja inscrita cuando se trataba de la unión de personas del mismo sexo, y la disposición negaba la pensión de supervivencia debido a ser pareja inscrita, se estaba discriminando directamente a causa de la orientación sexual (pár. 72). Alemania creó la posibilidad de establecerse como pareja inscrita para adaptar su ordenamiento jurídico y permitir la unión de personas del mismo sexo, habiendo optado por no abrir la institución del matrimonio, manteniéndose reservada por personas de distinto sexo. Por tanto, se creó un régimen diferente, que progresivamente fue asimilando las condiciones aplicables al matrimonio (pár. 67). Entonces, la pareja inscrita, sin ser idéntica al matrimonio, gozaba de una situación comparable en lo relativo a la prestación de supervivencia, que la disposición estatutaria controvertida negaba a las parejas inscritas (pág. 69-72).

En este caso se va más allá que en el caso *Dekker*, puesto que el nexo causal entre motivo protegido y disposición, práctica o criterio no era tan estrecho. No se trataba de una situación en la que solo podrían encontrarse las mujeres (embarazo), sino que era una disposición la que creaba la disparidad. Todos los beneficios que se asociaran al matrimonio no podían ser disfrutados por aquellas personas que no pudieran acceder a la institución del matrimonio, por lo que el matrimonio estaba indisociablemente ligado a la orientación sexual, entrando en un caso, entonces, de discriminación directa.

Esta interpretación extensiva de la existencia de un nexo causal sería interesante para abordar las situaciones en las que en un sistema algorítmico el indicador y el atributo protegido prácticamente se superponen. La jurisprudencia del TJUE considera una acción constitutiva de discriminación directa cuando el criterio formal en base al cual se toma una decisión es inseparable de la pertenencia al grupo desaventajado (Soriano Arnanz, 2021a: 16).

Se ha argumentado que bajo esta jurisprudencia, en la discriminación mediante sistemas algorítmicos, a pesar de la opacidad del algoritmo, podría determinarse que existe discriminación directa. Por ejemplo, cuando un sistema algorítmico sobre el que se desconoce el elemento o elementos que determinan la decisión tomada, rechaza a todas las mujeres y solo contrata a hombres (y todas las personas candidatas cumplen unos requisitos mínimos para ocupar los puestos de trabajo). Si en el resultado del algoritmo se evidencia un trato menos favorable para todos los miembros de un grupo determinado especialmente protegido, se podrá concluir que la decisión se ha tomado en base a la categoría especialmente protegida o en base a elementos de juicio indisociables de aquélla (Soriano Arnanz, 2022: 152, 153), en cuyo caso se podrá considerar la existencia de discriminación directa si se toma el algoritmo

en su totalidad como práctica discriminatoria. Estos casos se auguran, sin embargo, muy improbables.

III. DISCRIMINACIÓN ALGORÍTMICA: UNA FORMA INDIRECTA DE DISCRIMINACIÓN

La discriminación indirecta está orientada a realizar una valoración material de las desigualdades (Añón Roig, 2013: 153). Podría definirse como aquella situación en la cual una práctica que aparentemente es neutra acaba discriminando a personas dotadas de un atributo protegido[14]. Es una creación del Tribunal Supremo Federal de Estados Unidos de Norteamérica en la sentencia *Griggs contra Duke Power Company*, de 8 de marzo de 1971[15]. El tribunal estipuló que, aunque una práctica fuera neutral y no pretendiera discriminar, no se podía conservar si servía para mantener el *statu quo* de anteriores costumbres discriminatorias. En este sentido, Adams-Prassl *et al.* (2023: 154) afirman, en al ámbito de los sistemas algorítmicos, que los circuitos de retroalimentación[16] que se justifican a sí mismos básicamente lo que hacen es mantener un *statu quo* discriminatorio, como, por ejemplo, en el caso del exceso de vigilancia policial en ciertos barrios gracias a la policía predictiva.

Un ejemplo de discriminación indirecta a través de sistemas algorítmicos sería el siguiente: la persona (o personas) a cargo del sistema algorítmico, o el sistema algorítmico mediante el aprendizaje automático, establece que una de las características observables para clasificar como *buena* a cada persona candidata a un puesto de trabajo es *no haber interrumpido su carrera profesional en ningún momento*. El algoritmo, en este caso, estaría causando discriminación indirecta contra las mujeres, dado que a pesar de que *no haber interrumpido su carrera profesional en ningún momento* parece un criterio neutro, tiene una afectación más acusada en el caso de las mujeres que de los hombres. Esta afectación tiene su causa en que la maternidad y las tareas de cuidado son asignadas debido a un sistema de división de roles por razón

[14] Art. 2.2.b. de la Directiva 2000/43/CE; articulo 2.2.b. de la Directiva 2000/78/CE; articulo 2.b. de la Directiva 2004/113/CE i; articulo 2.1.b. de la Directiva 2006/54/CE.

[15] *Griggs v. Duke Power Co.*, 401 U.S. 424.

[16] Los circuitos de retroalimentación (en inglés *feedback loop*) se forman a través de perjuicios de asignación y representación a raíz de los sesgos y la discriminación existentes, creando dinámicas que se refuerzan y se trasladan a la toma de decisión mediante sistemas algorítmicos.

de género en el que las mujeres tienden a interrumpir su carrera profesional con mayor frecuencia que los hombres (Soriano Arnanz, 2022: 140).

Basándonos en la definición que se establece en las directivas 2000/43/CE, 2000/78/CE, 2004/113/CE, 2006/54/CE, podemos determinar cuatro requisitos para encontrarnos ante una situación de discriminación indirecta: un criterio, disposición o práctica aparentemente neutro, que sitúa a personas con una característica protegida por la legislación en desventaja particular con respecto a otras personas, sin que sea objetivamente justificable por una finalidad legítima y sin que los medios para conseguirla sean adecuados y necesarios.

1. ELEMENTOS QUE CONFORMAN LA DISCRIMINACIÓN INDIRECTA

En primer lugar, para apreciar la aparente neutralidad de un criterio o práctica hay que tener en cuenta el alcance de la ley. El tribunal debe cuestionarse si el caso puede integrarse en el campo de aplicación de la ley antidiscriminatoria europea que deben aplicar los Estados miembros. El TJUE defiende que los elementos que describen el alcance de manera positiva deben interpretarse ampliamente y, en cambio, los elementos que describen el alcance de forma negativa deben interpretarse restringidamente. Sin embargo, depende mucho de cómo se pone en práctica y la legislación de los Estados miembros debido a que estos tienen la capacidad de regular la proporcionalidad entre la seriedad de la intervención y la gravedad de las razones que la justifican (Tobler, 2008: 38).

En las decisiones automatizadas, la disposición, práctica o criterio aparentemente neutro pueden ser, por ejemplo, los algoritmos usados para tomar la decisión o las prácticas que se llevan a cabo sobre los datos que servirán de entrenamiento para el algoritmo (Allen y Masters, 2020: 592). Como la gran cantidad de datos tratados y la complejidad de los sistemas de tratamiento impiden identificar de forma clara cuáles son los elementos que condicionan los resultados en el sistema, apunta Hacker (2018: 1161, 1162) que será la totalidad del sistema el que deberá ser considerado como una disposición, práctica o criterio aparentemente neutro que genera resultados discriminatorios. De este modo, será más fácil abordar la causa de la discriminación, en lugar de buscar dentro del funcionamiento de un sistema inherentemente opaco[17].

[17] Se debe tener en cuenta que no todos los sistemas algorítmicos son inherentemente opacos. Los sistemas algorítmicos no predictivos permiten cierta comprensión del sistema al basarse en normas preestablecidas para llegar a un resultado concreto. En

En segundo lugar, para determinar que ha existido discriminación indirecta debe poder demostrarse que se ha sufrido una desventaja particular en relación con otras personas. En la discriminación directa, la persona demandante debe probar que fue tratada menos favorablemente en base a según un atributo protegido, probando un patrón ilegal, como, por ejemplo, que una compañía no contrata a personas de una etnia particular, a pesar de que una parte de la población sea de esa etnia. En cambio, la discriminación indirecta comporta un mayor uso de las estadísticas porque los atributos protegidos no son utilizados explícitamente (Makkonen, 2007: 31, 32). Como la discriminación indirecta se centra en el resultado más que en el trato, es necesario evidenciar el impacto de la medida que se pone en cuestión entre dos grupos, el de la presunta víctima y otro. En los sistemas algorítmicos, una vez que el algoritmo incorpora un sesgo que perjudica a las personas que pertenecen a un grupo desaventajado, salvo que incorpore instrucciones para corregirlo —lo que sucede de forma poco habitual—, se perpetúa durante todo el período de funcionamiento del sistema (Soriano Arnanz, 2021a: 21).

Si bien se ha recurrido a la prueba estadística tanto en casos presentados frente al TJUE (por ejemplo, asunto *Seymour-Smith y* Pérez[18] o *Violeta Villar Láiz*[19]) como en casos presentados delante del TEDH (por ejemplo, *D. H. y otros contra la República Checa*[20]), no se ha fijado por parte de ninguno de los dos tribunales un criterio único sobre qué porcentaje de personas negativamente afectadas por la disposición, criterio o práctica aparentemente neutra debe pertenecer al grupo desaventajado, ya que se tiende a analizar la situación de forma específica. Tampoco se ha establecido el porcentaje de personas con el mismo atributo que deben resultar negativamente afectadas en comparación con las demás que no comparten este atributo y que se encuentran en el grupo desaventajado (por ejemplo, entre las personas trabajadoras a tiempo parcial hay hombres y mujeres, aunque las mujeres constituyen un porcentaje más elevado) para considerar que se da un caso de discriminación indirecta (*ibid.*: 20).

---

cambio, se considera a los sistemas algorítmicos predictivos, que se basan en el análisis de datos, la correlación y la búsqueda de patrones, como opacos, ya que es muy complicado establecer y comprender en base a qué normas se ha llegado a un determinado resultado.

18 TJCE, sentencia de 9 de febrero de 1999, *Seymour-Smith y Pérez*, C-167/97, EU:C:1999:60.

19 TJUE, sentencia de 8 de mayo de 2019, *Violeta Villar Láiz*, C-161/18, EU:C:2019:382.

20 TEDH, sentencia de 13 de noviembre de 2007, 57325/00, *D. H. y otros c. República Checa*.

Sin embargo, a pesar de ser lógico pensar que se tendrán muchos datos estadísticos, dado que se analizan datos masivamente (Xenidis y Senden, 2020: 171), si la persona perjudicada no sabe explotarlos no podrán ser útiles a la hora de probar la discriminación. En Estados Unidos existe la norma de los cuatro quintos, que se basa en que si una tasa de selección por cualquier sexo, raza o grupo étnico es inferior a cuatro quintos de la tasa de selección del grupo con la mayor tasa, se debe considerar generalmente como evidencia de impacto adverso[21]. Esta regla adaptada a la discriminación algorítmica supondría reunir las estadísticas requeridas para demostrar que una política es indirectamente discriminatoria, proceso que es a menudo costoso y difícil (Tobler, 2022: 99), y que impone una carga técnica y financiera a la persona demandante (Páez, 2021: 24 , 25).

En tercer lugar, las justificaciones objetivas de discriminación indirecta no son *numerus clausus*, contrariamente a los casos de discriminación directa. Maliszewska-Nienartowicz (2014: 54) plantea que el hecho de que la discriminación indirecta normalmente no esté relacionada con la intención de trato desigual, ocasionando incluso un efecto discriminatorio accidental, supone una razón suficiente para una gama más amplia de posibles motivos de justificación que la discriminación directa. El TJUE ha tendido a aceptar todas aquellas justificaciones que no escondan el objetivo de discriminar de forma evidente (por ejemplo, caso *Achbita*[22] o *Bilka-Kaufhaus*). La justificación de la discriminación indirecta se basa en probar que la medida discriminatoria persigue una finalidad legítima, permitiendo al tribunal realizar un examen de proporcionalidad a través del cual se ponderen los intereses en juego (adecuación, necesidad y proporcionalidad en sentido estricto). No se pasará el examen de proporcionalidad si la finalidad puede alcanzarse a través de una medida diferente que sea menos discriminatoria o no discriminatoria. Siguiendo al Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial (2019: considerando 47), la IA debería ser usada para mejorar el bienestar colectivo e individual. Sin embargo, este concepto es bastante elástico y muchos algoritmos pueden ser aceptados como legítimos bajo esta base (Allen y Masters, 2020: 595). Además, la tendencia de las personas a confiar en la neutralidad de los sistemas automatizados aumenta la posibilidad de aceptación de los sesgos como necesarios, pudiendo llegar a concluir que superan el test

[21] Uniform Guidelines on Employment Selection Procedures adopted by the United States of North America Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice, 1978, par. 1607.4 (D).

[22] TJUE, sentencia de 14 de marzo de 2017, *Achbita*, C-157/15, EU:C:2017:203.

de proporcionalidad, ocasionando, por consiguiente, que los tribunales acaben aceptando decisiones automatizadas que perpetúen la situación de desventaja de grupos oprimidos (Soriano Arnanz, 2022: 162).

En efecto, desde una posición crítica en referencia a posibilidad de justificar objetivamente los efectos discriminatorios de un criterio, práctica o disposición, se plantea que el hecho de que haya grupos de población que hayan sido minusvalorados en la historia y, en consecuencia, no hayan formado parte en los procesos de formación de normas (por ejemplo mujeres), interpela directamente a la visión dominante de la discriminación indirecta que sostiene que esta puede justificarse «objetivamente» (Añón Roig, 2022: 42). También se argumenta que el hecho de que la discriminación en los tratos neutros resulte difusa y, por tanto, más complicada de identificar, no significa que deba poder justificarse «objetivamente» (Barrère, 2014: 522, 523).

En cuarto lugar, cuando nos refiramos a los sistemas algorítmicos, su adecuación como medida para conseguir una finalidad establecida seguramente dependerá de la existencia de otro algoritmo que pueda conseguir el mismo resultado sin tener unos efectos tan discriminatorios, ya que no existen otras formas actualmente de procesar la cantidad de datos que tratan los algoritmos en la misma cantidad de tiempo. Por tanto, teniendo en cuenta la opacidad y complejidad de la toma de decisiones mediante sistemas algorítmicos, es poco probable que se pueda analizar a) si todas las variables que se tienen en cuenta son necesarias para la toma de la decisión; b) si prescindiendo de ellas también se podría llegar al mismo resultado; c) si la base de datos utilizada es suficientemente representativa o contiene sesgos, o d) si el peso que se le da a cada variable es adecuado, resultando más probable que el test de adecuación se realice en relación al sistema en su conjunto y no por separado (Hacker, 2018: 1161, 1162). En cualquier caso, el examen de adecuación parece difícil dada la naturaleza opaca de los algoritmos que se emplean (Páez, 2021: 20).

La parte demandada, en teoría, deberá acreditar que valoró el uso de otros sistemas y bases de datos y que no existía un algoritmo menos discriminatorio ni una base de datos más completa y menos sesgada; sin embargo, se deberá tener en cuenta también la capacidad técnica y económica de la parte demandada (Soriano Arnanz, 2021a: 25), lo que establece un amplio margen de apreciación. La persona demandante podrá tener éxito en su pretensión si muestra que la persona demandada podría haber utilizado un sistema alternativo con resultados menos discriminatorios. También se ha plateado que el establecimiento de un sistema de mejores técnicas disponibles facilitaría el enjuiciamiento de decisiones tomadas mediante sistemas algorítmicos, ya que en caso de que se estuviera planteando la existencia de discriminación algorítmica, el tribunal tendría la posibilidad de determinar

si existe otro sistema algorítmico que, sin ser más costoso ni menos preciso, produce resultados menos discriminatorios en el mismo contexto (Soriano Arnanz, 2021b: 120-121).

En definitiva, el principal problema de la vertiente indirecta de la discriminación algorítmica no será probar que la tipología de discriminación ha sido indirecta, sino conseguir elementos probatorios que constaten la existencia de discriminación. Aun así, será un paso menos que evidenciar discriminación directa, en la que, seguramente, una vez constatada la discriminación, habrá que probar que esta ha sido directa.

2. APUNTES DE LA JURISPRUDENCIA DE TJUE EN RELACIÓN CON LA DISCRIMINACIÓN INDIRECTA

En los asuntos *Jenkins* y *Bilka-Kaufhaus* del TJUE relativos a la discriminación retributiva por razón de sexo se establece que las personas trabajadoras a tiempo parcial no pueden ser tratadas de forma menos favorable que las personas trabajadoras a tiempo completo si se trata de un medio indirecto para discriminar por razón de sexo. Se argumenta en estos casos que, teniendo en cuenta las dificultades encontradas por las trabajadoras para trabajar a jornada completa y las presiones socioculturales que pesan sobre ellas dadas las cargas familiares, el grupo de personas trabajadoras a tiempo parcial está integrado de forma exclusiva o preponderante por mujeres.

Aplicada a los sistemas algorítmicos, esta jurisprudencia podría tenerse en cuenta en relación con los criterios que establecen las diferencias de remuneración en las plataformas colaborativas. No son criterios que se basan directamente en el sexo de las personas, pero están relacionados con la forma en que se construyen los roles de género, es decir, el hecho que las mujeres sigan cargando con una mayor parte de la responsabilidad de cuidado de familiares continúa afectando a sus carreras y su posición económica (Arabadjieva y Zwysen, 2022: 7).

Soriano Arnanz (2022: 162) argumenta que aplicándose la jurisprudencia del caso *Bilka-Kaufhaus*, la forma en que se calculan, a través de sistemas algorítmicos, los salarios en las plataformas de transporte colaborativo, perjudica de forma desproporcionada a las mujeres. No obstante, parece difícil aislar concretamente la causa de la diferencia de salarios. Puede suceder, por ejemplo, que se trate de un sistema retributivo caracterizado por una falta absoluta de transparencia, de modo que solo pueda llegar a la conclusión de que los hombres cobran más que las mujeres. Sin embargo, entonces, el TJUE se inclina por atribuir la carga de la prueba a la persona empresaria, que debe acreditar el carácter neutro del sistema retributivo aplicado (Cabeza Pereiro, 2011: 91). También se constata así en la Directiva

(UE) 2023/970, por la que se refuerza la aplicación del principio de igualdad de retribución entre hombres y mujeres por un mismo trabajo o un trabajo de igual valor a través de medidas de transparencia retributiva y de mecanismos para su cumplimiento, que en su art. 18.2 establece que «cuando el empleador no haya cumplido las obligaciones de trasparencia retributiva […] [le] correspond[erá] […] en relación con una presunta discriminación directa o indirecta en relación con la retribución, demostrar que no se ha producido tal discriminación».

Gerards y Xenidis (2021: 71-73), defienden que el concepto de discriminación indirecta tal y como interpretado por el TJUE, a pesar de la dificultad de obtener prueba y de la posibilidad de plantear justificaciones objetivas, es capaz de abordar mejor las situaciones de discriminación a raíz de indicadores cuando se debate si la vinculación entre el indicador y el atributo protegido no es suficientemente evidente como para considerarse discriminación directa. Sería el caso de situaciones en las que un algoritmo emplea los datos de residencia y código postal para inferir el origen étnico de las personas y discriminarlas por ello. También en caso de que no se pueda establecer la relación entre el indicador y atributo protegido, es decir, cuando los datos empleados para programar o entrenar el sistema algorítmico reflejan sesgos y estereotipos que se han cristalizado en patrones de desigualdad y tratan de forma más desfavorable a personas pertenecientes a un colectivo social determinado, pero no se puede establecer exactamente por qué. En este caso, podrían estar produciéndose discriminaciones directas (trato desigual debido a la pertenencia a un grupo protegido), pero dada la opacidad de los sistemas algorítmicos parecen imposibles probar que así ha sido.

Sin embargo, Sáez Lara (2020: 49) cuestiona la fácil equiparación de la discriminación algorítmica con un supuesto de discriminación indirecta (en el ámbito laboral pero equiparable a otros ámbitos), ya que cuando hay que hacer frente a una decisión automatizada, el problema desborda o supera el supuesto de hecho dada su complejidad, y sugiere una nueva ampliación de la tutela antidiscriminatoria. Añón Roig (2022: 43, 44) plantea la reconducción de la discriminación algorítmica hacia modalidades distintas a la discriminación indirecta (específicamente la discriminación por asociación y la discriminación interseccional), dado que considera que estas categorías pueden superar algunas deficiencias del concepto de discriminación indirecta.

IV. ¿CUÁNDO LA DISCRIMINACIÓN ALGORÍTMICA ES DIRECTA Y CUÁNDO ES INDIRECTA?

En este apartado nos preguntamos: ¿cómo se adapta la discriminación mediante sistemas algorítmicos a los conceptos de discriminación directa e indirecta?

Rivas Vallejo (2022: 67, 68) pone el ejemplo de los sistemas algorítmicos de contratación. Estaríamos ante discriminación directa si la empresa (la que crea el sistema para utilizarlo directamente o la que lo comercializa para otras empresas) traslada o introduce parámetros de sesgo en el algoritmo (este se construye en base a órdenes que persiguen un resultado concreto y ese define por quien ordena su programación).

En cambio, estaríamos ante discriminación indirecta en el caso de un algoritmo de aprendizaje automático que funciona de forma opaca y simplemente valora todos los elementos que han puesto a su disposición para llegar a la *mejor decisión*, que resulta sesgada, pero de forma ajena a una motivación de quien ha programado el algoritmo. Si, igualmente, la empresa toma de referencia un patrón histórico discriminatorio para llegar a la *mejor decisión*, es decir, los antecedentes de sesgo de la propia empresa, estaría llevando a cabo un trato abiertamente desfavorable hacia aquellas personas dotadas de un atributo protegido (por ejemplo, sexo u origen étnico) que se han visto perjudicadas por la política de contratación en el pasado, por lo que la discriminación también sería directa. La autora hace este planteamiento para concluir que en el ámbito laboral la persona empresaria sigue siendo responsable de la acción discriminatoria, siendo determinante la actitud de la empresa frente al posible efecto perverso del algoritmo, es decir, su consentimiento para validar el sesgo discriminatorio que reproduce, confirmando la propuesta del modelo matemático.

Por tanto, se presentan diferentes casos:

1) Si se han dado al sistema algorítmico directrices concretas y evidentes para tratar desfavorablemente a un determinado grupo de población que comparte un atributo protegido, estaremos ante discriminación directa.
2) Si el sistema algorítmico es de aprendizaje automático y se base en los patrones, tendencias o reglas sociales que ha determinado en los datos, y simplemente valora todos los elementos que se han puesto a su disposición para tomar la *mejor decisión*, estaremos ante un caso de discriminación indirecta, dado que el algoritmo habrá tenido en cuenta otros factores que no son únicamente el atributo protegido (y que no son equivalentes al mismo) para tomar la decisión, pero el

resultado habrá sido más desfavorable para un grupo de personas que comparten un atributo protegido determinado. Sin embargo, estos otros factores no pueden ser *equivalentes al atributo protegido*, puesto que, si son equivalentes, podríamos estar ante un caso de discriminación directa. Nos podríamos preguntar aquí si en todos los casos que se utilicen sistemas algorítmicos que incluyan datos de comportamiento habrá discriminación dado que los resultados mostrarán el estado actual del mundo, fundamentalmente desigual. Así que, todo el mundo que utilice sistemas algorítmicos de toma de decisión utilizará bases de datos discriminatorias (que reflejan el estado pasado o actual del mundo) que darán resultados indefectiblemente indisociables del trato discriminatorio que reciben las personas que ostentan atributos protegidos (por tanto en cualquier caso se tratará de discriminación directa al ser conscientes, las personas que utilicen sistemas algorítmicos para la toma de decisiones, que con su uso están perpetuando dinámicas de desigualdad existentes en la sociedad).

3) Si el sistema algorítmico es de aprendizaje automático y para llegar a la «mejor decisión» utiliza datos que vienen dados en base a un patrón histórico discriminatorio (que se puede concretar en los datos históricos de contratación de una empresa, por ejemplo), entonces se trata de discriminación directa, dado que de entrada quien emplea el algoritmo ya sabrá que los resultados que revele serán discriminatorios y que el trato ya era desigual desde el principio.

Parece ser, entonces, que en el estado actual de la cuestión la principal característica que distingue la discriminación directa de la discriminación indirecta es el conocimiento por parte de la persona que toma la decisión sobre la relación entre el atributo protegido y el resultado.

Hay que tener en cuenta en la distinción de discriminación directa e indirecta planteada por la jurisprudencia del TJUE que toda la construcción de los conceptos está creada basándose en la discriminación por un atributo protegido. Por tanto, no se aborda una posible multiplicidad de atributos protegidos que intervengan en la decisión.

Según las distinciones hechas por Rivas Vallejo, podemos identificar dos elementos que dificultan la distinción entre discriminación directa e indirecta (es decir, la distinción entre si había o no conocimiento sobre la relación entre el atributo protegido y el resultado). En primer lugar, la posibilidad de hacer uso de indicadores muy cercanos al atributo protegido que puedan ser, a fin de cuentas, equivalentes al atributo protegido, pero que no se pueda determinar que son equivalentes. En segundo lugar, el carácter subrepticio de la

discriminación algorítmica, que puede causar que, a pesar de las discriminaciones sean directas, sea imposible probarlo.

1. FACTORES EQUIVALENTES AL ATRIBUTO PROTEGIDO

El funcionamiento de los sistemas algorítmicos y su opacidad generan, entre la distinción tradicionalmente establecida de discriminación directa y discriminación indirecta, un reto de clasificación, dado que será cada vez más difícil identificar tratamiento diferencial basado en atributos protegidos en el contexto de las operaciones algorítmicas (Gerards y Xenidis, 2021: 76). En este sentido, concluyen Barocas y Selbst (2016: 713) que la discriminación directa y la discriminación indirecta se convierten esencialmente en lo mismo desde una perspectiva probatoria.

En el caso *Szpital Kliniczny* el tribunal establece que existe discriminación directa por razón de discapacidad cuando el tratamiento desfavorable se basa en un criterio inextricablemente vinculado a la discapacidad (pár. 48). Usualmente existe una clara superposición entre el motivo protegido y sus indicadores, es decir, que el uso del indicador cubre casi exactamente el mismo grupo de personas que cubriría el motivo protegido (Gerards y Xenidis, 2021: 76). En el caso de los sistemas algorítmicos, se debería debatir hasta qué punto ciertos indicadores se podrían considerar como atributos protegidos al revelar la misma información.

Si se les da una interpretación extensiva, los indicadores más relevantes también deberían estar cubiertos por la protección de discriminación directa, puesto que, a pesar de no ser exactamente el atributo, se produce una superposición del significado casi completa. El ejemplo clásico es la discriminación basada en el embarazo, que está claramente relacionada con la discriminación por razón de sexo, ya que embarazo es un indicador de mujer. Como hemos establecido, el TJUE ha tratado la discriminación por embarazo como discriminación directa (caso *Dekker*). Otro buen ejemplo es tener un pasaporte extranjero, que es un claro indicador de tener una nacionalidad diferente.

Por el contrario, si se entiende el atributo protegido de forma restrictiva, solo cuando se discrimine basándose en este estaremos hablando de discriminación directa, y toda la discriminación que provenga de los indicadores o correlaciones sobre este atributo se considerará discriminación indirecta. Por tanto, un reto importante para dilucidar la distinción entre discriminación directa e indirecta cuando la discriminación se produce a raíz del uso de sistemas algorítmicos es descubrir qué indicadores llegan a tener un grado tan grande de superposición con el motivo protegido que pueden ser vistos como la misma cosa (Adams-Prassl *et al.*, 2023: 171). Debe cuestionarse entonces si es necesario un 100 % de superposición o estadísticamente también se admite

un indicador que tenga un 90 % o 80 % de superposición con un atributo protegido; y si este tratamiento desigual puede seguirse considerando discriminación directa o debe considerarse indirecta (Gerards y Xenidis, 2021: 64).

Probar un tratamiento diferencial basado en atributos protegidos o indicadores puede resultar muy complicado y puede ocasionar que al final, al no localizar si la discriminación proviene directamente de un atributo sensible (sexo, origen étnico, etc.) o de un indicador (embarazo, pasaporte extranjero , etc.), se acabe sin poder determinar exactamente si se trata de discriminación directa o indirecta, haciendo que al final sea la discriminación indirecta la que actúe como refugio para captar la discriminación algorítmica. Esto puede implicar, como ya hemos dicho, la reducción de la seguridad jurídica si supone que podrá aplicarse el sistema de justificaciones objetivas abiertas en vez del conjunto *numerus clausus* de justificaciones disponibles para la discriminación directa (Gerards y Xenidis, 2021: 9). Para detectar la discriminación directa en los sistemas algorítmicos se ha propuesto que pueda experimentarse con los sistemas (Soriano Arnanz, 2021c: 81), y así tener una herramienta para intentar evitar que la discriminación indirecta actúe como refugio de la discriminación algorítmica.

2. OTRAS CAUSAS QUE MOTIVAN QUE LA DISCRIMINACIÓN ALGORÍTMICA RECAIGA EN LA DISCRIMINACIÓN INDIRECTA

El tratamiento de los datos y su categorización por parte de los algoritmos puede no ser cognoscible por la persona y, por tanto, que sea imposible determinar dónde está la discriminación directa (Leese, 2014: 494). Además, el uso de variables y categorización de datos en los algoritmos de aprendizaje automático está en constante evolución mientras el algoritmo aprende. Como estas categorías no son estáticas, puede ser difícil saber si están relacionadas con los atributos protegidos. Habría que mirar el modelo algorítmico y la forma en que el modelo estadístico usado trata los datos disponibles a lo largo del tiempo para encontrar el tratamiento desfavorable, una situación que puede ser difícil de cumplir dados los problemas de accesibilidad (Gerards y Xenidis, 2021: 69).

Dada la prohibición de utilizar atributos protegidos en la toma de decisiones, es poco probable el empleo de estos en los sistemas algorítmicos. Además, las personas que desarrollen los sistemas algorítmicos podrían no introducir en el algoritmo los datos relacionados directamente con atributos protegidos (Schreurs *et al.*, 2008: 260) para no ocasionar una disminución de la precisión algorítmica, que se incrementa cuantas más correlaciones diversas se pueden hacer (Hacker, 2018: 1152, 1161). Igualmente, en el caso de incluir voluntariamente los atributos protegidos, la normativa antidiscriminación es

fácil de eludir haciendo uso de indicadores. Probar que el disfraz de los indicadores es intencionado es difícil, si no imposible, ya que quien los ha utilizado siempre puede argumentar que no era consciente de la discriminación que se estaba creando. De esta forma, estos métodos de discriminación intencional parecerán, a todos los efectos, idénticos a la discriminación no intencional que puede resultar del análisis de datos (Páez, 2021: 26).

Se añade que para un tribunal es difícil —si no imposible— hacer una retrospectiva y reconstruir las numerosas evaluaciones sesgadas y percepciones que han resultado en la decisión adversa (Barocas y Selbst, 2016: 698), basada en sesgos implícitos e inconscientes que se dan cuando alguna persona ha internalizado un estereotipo social de tal forma que aplica a una persona un tratamiento ostensiblemente más desfavorable que a otra. Los sistemas algorítmicos, al aplicar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de datos en un determinado contexto, perpetúan acciones discriminatorias no intencionales (Barocas y Selbst, 2016: 693) —más fáciles de pasar por alto— que las personas hemos ido adoptando debido a vivir en una sociedad desigual y marcada por las discriminaciones sistémicas y estructurales (típicamente encuadradas en el concepto de discriminación indirecta).

Por último, se ha subrayado que cuando las obligaciones legales aumentan, la discriminación directa es proclive a disminuir en la fase de desarrollo de los algoritmos (Hacker, 2018: 1152).

V. CAMBIAR DE PERSPECTIVA: ¿Y SI LA DISCRIMINACIÓN ALGORÍTMICA SE ESTABLECIESE BASÁNDOSE EN LA PRECISIÓN DEL SISTEMA SEGÚN LOS GRUPOS DE POBLACIÓN?

Si la discriminación causada sin que se vean implicados sistemas algorítmicos ya es difícil de probar, con la introducción de los sistemas algorítmicos en la ecuación puede resultar prácticamente imposible, sobre todo teniendo en cuenta que las personas a las que se aplicará el sistema muy probablemente no tendrán conocimientos sobre su funcionamiento, no tendrán acceso a información ininteligible y coherente como para crearse un criterio propio y en muchos casos incluso ni serán conscientes de que están sufriendo discriminación[23]. Si bien teóricamente podemos especular sobre en cuál de las dos

[23] Hay que tener en cuenta, sin embargo, que el reglamento europeo de la IA sí que establece ciertos requisitos que deben cumplir los sistemas de IA de riesgo alto para mitigar los efectos discriminatorios de estos sistemas o mejorar su comprensibilidad, como, por ejemplo, la posibilidad de pedir una explicación en relación con una deci-

categorías (directa e indirecta) se podrán integrar los supuestos de discriminación algorítmica, en cualquier caso será difícil de probar que ha ocurrido. En el caso de la discriminación directa porque será muy difícil probar que el trato perjudicial es a causa de uno (o varios) atributo(s) protegido(s) concreto(s), y en la discriminación indirecta porque en relación con los efectos discriminatorios, al estar las personas perjudicadas plurilocalizadas sin existir grupos concretos, será muy difícil tanto ser consciente de la discriminación como establecer grupos de comparación.

Por consiguiente, creemos pertinente considerar si estas categorías (directa/indirecta) deberían verse sustituidas por algún criterio más técnico que facilitase la presentación de pruebas relativas a la existencia de discriminación, adaptando al entorno digital en constante evolución los mecanismos de protección del principio fundamental de no discriminación (Sánchez Hernández, 2024: 301). Ahondamos en nuestro argumento sobre el necesario replanteamiento de los mecanismos de protección frente a la discriminación algorítmica siguiendo a Barrère Unzueta (1997: 74), que ya en 1997 planteaba que la clasificación de directa o indirecta no hacía tanto referencia a la discriminación estrictamente, sino al proceso para identificarla o detectarla. Pues bien, ponemos en duda la adecuación de los procesos de identificación y detección establecidos para abordar la discriminación ocasionada por unos sistemas que disponen de unos medios para causar (amplificar, perpetuar, sutilizar, enmascarar) la discriminación de los que no se disponía hasta ahora.

En consecuencia, nos parece relevante plantear herramientas específicas de detección e identificación de la discriminación algorítmica como, por ejemplo, la normativización de un cribado de precisión exigible a cada sistema de IA que tenga incidencia en el ámbito social, teniendo en cuenta la perspectiva interseccional. Es decir, asumiendo que la regulación antidiscriminación actual no tiene como objetivo ser una herramienta transformadora de las estructuras sociales de opresión/subordinación (porque si así fuera deberían o bien prohibirse todos aquellos sistemas algorítmicos aplicados a contextos sociales en tanto que perpetúan a gran escala las dinámicas sociales desiguales, o bien exigir que los sistemas algorítmicos integrasen medidas de acción positiva para contrarrestar la subordinación de ciertos grupos sociales), sino que dentro de las estructuras ya creadas, tiene vocación de corregir evidencias

sión tomada por un sistema de IA (art. 86), la realización de una evaluación de impacto en los derechos fundamentales (art. 27) o la habilitación excepcional del tratamiento de categorías especiales de datos para detectar y corregir sesgos algorítmicos que puedan contribuir a crear efectos discriminatorios (art. 10.5).

flagrantes puntuales de desventaja para ciertos grupos de población, por lo que podría ser pertinente asegurar que la precisión del sistema algorítmico para todos los grupos de población concretos (incluyendo categorías de datos combinados) no supere cierta tasa de error y centrarse en que, al menos, los sistemas algorítmicos magnifiquen lo menos posible la discriminación ya existente. Esta parece una apuesta coherente y factible dentro de los límites de la normativa antidiscriminación actual, concretando una normativa que exija a los sistemas algorítmicos la consecución de unos estándares concretos en relación con la posibilidad de crear resultados discriminatorios.

Entonces, la apreciación de discriminación dependería de si el sistema demuestra o no que alcanza los estándares de precisión exigidos y a partir de estos establecer controles *ex ante* y *ex post*. Por ejemplo, la normativa puede establecer que no se pueden utilizar sistemas en los que las tasas de error por todos los grupos —teniendo en cuenta también la interseccionalidad (por ejemplo, un grupo de población que sean mujeres, afroamericanas de más de sesenta años)— supere el 1 %. En el mismo sentido, pero estableciendo otro sistema de cribado, la normativa puede exigir que el sistema algorítmico sea un 95 % preciso para todos los grupos de población (incluyendo categorías protegidas combinadas). Buolamwini y Gebru (2018: 77) presentan un método para evaluar el sesgo presente en conjuntos de datos y algoritmos de análisis facial automatizados respecto a los subgrupos fenotípicos. Se evalúan tres sistemas comerciales de clasificación según el género, que muestran que las mujeres negras son el grupo peor clasificado (con tasas de error hasta el 34,7 %), cuando, por el contrario, la máxima tasa de error para los hombres blancos es del 0,8 %. Entonces, por ejemplo, al existir disparidades sustanciales en la exactitud del sistema según los rasgos físicos de las personas sobre las que se aplica, estos no podrían emplearse. En cualquier caso, queremos remarcar que alcanzar ciertos estándares de precisión no exime de cuestionarse la adecuación del sistema, por ejemplo, en los casos de sistemas de vigilancia masiva que funcionan adecuadamente y sin errores y que igualmente producen efectos indeseables y reproducen patrones de discriminación (Katell *et al.*, 2020: 52).

Tanto en lo referente al control *ex ante*, como al control *ex post*, lo que debería comprobarse, en consecuencia, es que el sistema algorítmico empleado alcanza estos estándares, de modo que ambos tipos de control deberían pivotar sobre esta norma. El control *ex post* (que debería ser hecho por terceras partes independientes, no por la propia empresa con intención de comercializarlo o que lo está comercializando), en todo caso debería ser periódico y comprobar si con el paso del tiempo el algoritmo, a pesar de cambiar, sigue siendo tan preciso como exige la normativa. Si en sede judicial se demostrase que el sistema algorítmico o bien tiene una tasa de error más alta que aquella

permitida, o bien es tan inescrutable que no puede demostrarse cuáles son sus tasas de error, debería considerarse discriminatorio. En este sentido, sería el colectivo que se viese discriminado por el sistema algorítmico el que sería el sujeto de derecho y susceptible de protección (Sánchez Hernández, 2024: 288).

En definitiva, si tenemos en cuenta que tanto las personas como los sistemas algorítmicos no pueden no discriminar al cien por cien[24], lo que aquí proponemos es comprender que el alcance de la normativa es limitado y que para poder, al menos, plantear una regulación que pueda ser efectiva en caso de que en la discriminación se vean implicados sistemas algorítmicos, hay que entender que, a pesar de que la causa de la discriminación es la misma (una sociedad fundamentalmente desigual), los medios que emplea el derecho para captar la discriminación empleados hasta ahora (dicotomía directa/indirecta) se muestran inefectivos y que ante una sofisticación tecnológica debe responderse también con una reevaluación de la normativa.

VI. REFLEXIONES FINALES

La distinción entre discriminación directa e indirecta en el contexto de la discriminación algorítmica no parece operativa. El esquema que sigue el TJUE para distinguir entre discriminación directa y discriminación indirecta parece tener una relación clara al distinguir si la persona encargada de tomar la decisión era consciente o podría haber sido consciente de la relación entre el atributo protegido y el resultado discriminatorio.

En el caso de sistemas algorítmicos parece imposible poder dilucidar si existía consciencia sobre la relación entre el uso del sistema algorítmico y el resultado discriminatorio. Se pueden dar las siguientes situaciones: si el algoritmo tiene directrices concretas para tratar desfavorablemente a un grupo determinado de población que comparte un atributo protegido, si el sistema algorítmico emplea datos que vienen dados en base por un patrón histórico de discriminación, si utiliza atributos equivalentes a las categorías protegidas o que son inseparables o inextricablemente vinculados a las categorías protegidas, estamos ante discriminación directa. Sin embargo, será prácticamente

[24] La normativa antidiscriminatoria no exige a las personas no discriminar, sino que en su interacción con terceros y en la toma de decisiones que les afectan respeten los límites establecidos. Del mismo modo, los sistemas algorítmicos usados en el contexto social, en su interacción con terceros y en la toma de decisiones que les afectan, deberían evidenciar que, al menos técnicamente, actúan dentro de unos límites concretos.

imposible probar el conocimiento por parte de las personas que han empleado el sistema algorítmico de la relación o la implicación de los atributos protegidos en la obtención de un resultado discriminatorio por parte del sistema, por lo que, dada la opacidad inherente en el funcionamiento del sistema algorítmico, lo más probable será considerar que se ha producido discriminación indirecta a la vista del resultado discriminatorio.

Consecuentemente, deben tenerse en cuenta varias cuestiones. En primer lugar, confiar en que la persona que se ha visto discriminada pueda demostrar —después de haber demostrado que ha sufrido discriminación— que esta ha sido directa, ya que la persona que ha empleado el sistema de toma de decisión era consciente de que los resultados serían discriminatorios o por haberse utilizado atributos directamente protegidos o equivalentes, inseparables o inextricablemente vinculados, no parece muy probable. Si no se puede demostrar esta relación, la discriminación causada por decisiones tomadas mediante sistemas algorítmicos recaerá en la doctrina de la discriminación indirecta, que permite justificaciones objetivas (concepto que también es muy discutible desde la perspectiva antisubordinatoria, dada la imposibilidad de evaluar la objetividad en una sociedad fundamentalmente desigual) y presenta dificultades de prueba, sobre todo en relación con el establecimiento tanto de un grupo desaventajado como de un grupo comparador, ya que raramente se formará un grupo con suficiente consistencia teniendo en cuenta que las personas se encontrarán aisladas y los sistemas algorítmicos pueden excluir a personas del resultado directamente sin que ellas sean conscientes y discriminar de manera sutil y no evidente en base a múltiples factores (atomizando tanto a los grupos perjudicados como a los comparadores).

Aunque, según el TJUE, probar intencionalidad dejó de ser relevante para apreciar discriminación, sí parece seguir siendo relevante para que el tribunal aprecie discriminación directa o indirecta. Es decir, la lectura de los casos en los que el TJUE da criterios interpretativos para distinguir la discriminación directa de la indirecta trasluce que se busca dilucidar si el criterio controvertido se ha empleado específicamente para discriminar (aunque se haya usado con subterfugios —atributo indisociable o inextricablemente vinculado—) o ha terminado discriminando, pero era difícilmente probable que se pudiera saber por adelantado.

Debe tenerse en cuenta que la distinción fue pensada en un momento en que era factible (a pesar de ser complicado) probar que los resultados discriminatorios se habían dado en un contexto en el cual quien tomaba la decisión era consciente o podía ser consciente que la acción, criterio o práctica empleado generaba resultados discriminatorios (discriminación directa, por ejemplo, declaración pública en el caso *Feryn, Accept* o *Rete Lenford*) o no (discriminación indirecta). Sin embargo, en el caso de la discriminación algorítmica,

esta distinción puede ser imposible de hacer. Desde la perspectiva antisubordinatoria es una distinción poco relevante. Está planteada desde la perspectiva en la que no existen sistemas de opresión en los que todos participamos, sino que hay ciertas personas que no actúan como sería deseable y llevan a cabo acciones discriminatorias (siendo conscientes —discriminación directa—, o siendo inconscientes —discriminación indirecta—).

Teniendo en cuenta que efectuar esta distinción cuando están implicados sistemas algorítmicos es muy complicado, no parece configurar un recurso lógico para encuadrar la discriminación algorítmica. Por eso, planteamos la posibilidad de establecer, en el caso de que se usen sistemas algorítmicos para la toma de decisiones en el ámbito social y en vista de las limitaciones que plantea la normativa antidiscriminatoria, estándares de precisión según los grupos determinados de población a los que se aplicará el sistema algorítmico (incluyendo categorías combinadas de atributos protegidos) para asegurar que, al menos, el sistema no magnifica la discriminación ya existente en la sociedad.

Bibliografía

Adams-Prassl, J., Binns, R. y Kelly-Lyth, A. (2023). Directly Discriminatory Algorithms. *Modern Law Review,* 86 (1), 144-175. Disponible en: https://doi.org/10.1111/1468-2230.12759.

Agencia Europea de Derechos Fundamentales y Consejo de Europa. (2019). *Manual de legislación europea contra la discriminación*. Luxemburgo: Oficina de Publicaciones de la Unión Europea.

Allhutter, D., Cech, F., Fischer, F., Grill, G. y Mager, A. (2020). Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective. *Frontiers Big Data*, 3, 1-17. Disponible en: https://doi.org/10.3389/fdata.2020.00005.

Allen, R. y Masters, D. (2020). Artificial Intelligence: The right to protection from discrimination caused by algorithms, machine learning and automated decision-making. *ERA Forum,* 20 (4), 585-598. Disponible en: https://doi.org/10.1007/s12027-019-00582-w

Añón Roig, M. J. (2013). Principio antidiscriminatorio y determinación de la desventaja. *Isonomía: Revista de Teoria y Filosofia del Derecho,* 39, 127-157. Disponible en: https://doi.org/10.5347/39.2013.109.

Añón Roig, M. J. (2022). Desigualdades algorítmicas: conductas de alto riesgo para los derechos humanos. *Derechos y Libertades,* 47 (2), 17-49. Disponible en: https://doi.org/10.20318/dyl.2022.6872.

Arabadjieva K. y Zwysen, W. (2022). *Gender inequality in performance-related pay: A gap in the EU equal pay agenda*. ETUI Research Paper-Policy Brief. Disponible en: https://doi.org/10.2139/ssrn.4031499.

Barocas, S. y Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review*, 104, 671-732.Disponible en: https://doi.org/10.2139/ssrn.2477899.

Barrère, M. A. (1997). *Discriminación, derecho antidiscriminatorio y acción positiva en favor de las mujeres.* Madrid: Civitas.

Barrère, M. A. (2008*).* Iusfeminismo y derecho antidiscriminatorio: hacia la igualdad por la discriminación. En R. M. Mestre (coord.). *Mujeres, derechos y ciudadanías* (pp. 45-71). Valencia: Tirant lo Blanch. Disponible en: https://doi.org/10.47623/ivap-rvap.99.100.2014.021.

Barrère, M. A. (2014). La igualdad de género desde el activismo de las profesiones jurídicas. *Revista Vasca de Administración Pública,* 1 (99-100), 513-528.

Benedi Lahuerta, S. (2016). Ethnic discrimination, discrimination by association and the Roma community: CHEZ. *Common Market Law Review*, 53 (3), 797-818. Disponible en: https://doi.org/10.54648/COLA2016066.

Buolamwini, J. y Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. En *Proceedings of the 1st Conference on Fairness, Accountability, and Transparency,* 81 (pp. 77-91).

Cabeza Pereiro, J. (2011). La discriminación retributiva por razón de sexo como paradigma de la discriminación sistèmica. *Lan Harremanak-Revista de Relaciones Laborales*, 25, 79-98.

Foucault, M. (1979). *Microfísica del poder*. Madrid: Ediciones la Piqueta.

Fredman, S. (2011). *Discrimination Law*. Oxford: Oxford University Press.

Gerards, J. y Xenidis, R. (2021). *Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non-Discrimination Law*. European Commission. Disponible en: https://is.gd/uc48Bp.

Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial. (2019). *Directrices éticas para una IA fiable*. Bruselas: Comisión Europea.

Hacker, P. (2018). Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law. *Common Market Law Review,* 55 (4), 1143-1185. Disponible en: https://doi.org/10.54648/COLA2018095.

Katell, M., Young, M., Dailey, D., Herman, B., Guetler, V., Tam, A., Binz, C., Raz, D., y Krafft, P. (2020). Toward situated interventions for algorithmic equity: Lessons from the field. En M. Hildebrant, C. Castillo (eds.). *FAT'20. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (Barcelona, January 27-30, 2020)* (pp. 45-55). Disponible en: https://doi.org/10.1145/3351095.3372874.

Leese, M. (2014). The New Profiling: Algorithms, Black Boxes, and the Failure of AntiDiscriminatory Safeguards in the European Union. *Security Dialogue,* 45 (5), 494-511. Disponible en: https://doi.org/10.1177/0967010614544204.

Makkonen, T. (2007). *Measuring Discrimination: Data Collection and EU Equality Law*. European Network of Legal Experts in the Non-discrimination field. Disponible en: https://is.gd/ap4ElL.

Maliszewska-Nienartowicz, J. (2014). Direct and indirect discrimination in the European Union Law. How to draw a dividing line? *International Journal of Social Sciences,* 3 (1) 41-55.

Páez, A. (2021). Negligent algorithmic discrimination. *Law and Contemporary Problems,* 84 (3), 19-33. Disponible en: https://doi.org/10.2139/ssrn.3765778.

Rey Martínez, F. (2019). *Derecho antidiscriminatorio*. Cizur Menor: Aranzadi.

Rivas Vallejo, P. (2022). Sesgos de género en el uso de inteligencia artificial para la gestión de las relaciones laborales: análisis desde el derecho antidiscriminatorio. *e-Revista Internacional de la Protección Social (e-RIPS),* 8 (1), 52-83.

Sáez Lara, C. (2020). El algoritmo como protagonista de la relación laboral. Un análisis desde la perspectiva de la prohibición de discriminación. *Temas Laborales: Revista Andaluza de Trabajo y Bienestar Social,* 155, 41-60.

Sánchez Hernández, J. (2024). Posthumanismo, tecnología y evolución generacional de los derechos humanos: hacia la protección contra la discriminación algorítmica y el uso transparente y responsable de la IA. *Revista general de derecho constitucional*, 40, 283-316.

Schreurs, W., Hildebrandt, M., Kindt, E. y Vanfleteren, M. (2008). Cogitas, Ergo Sum. The role of data protection Law and non-discrimination law in group profiling in the private sector. En M. Hidelbrandt, y S. Gutwirth (eds.). *Profiling the European citizen: Cross-Disciplinary Perspectives* (pp. 241-270). Berlin: Springer Dordrecht. Disponible en: https://doi.org/10.1007/978-1-4020-6914-7_13.

Soriano Arnanz, A. (2021a). Decisiones automatizadas y discriminación: aproximación y propuestas generales. *Revista General de Derecho Administrativo*, 56, 1-45.

Soriano Arnanz, A. (2021b). Decisiones automatizadas: problemas y soluciones jurídicas. Más allá de la protección de datos. *Revista de Derecho Público: Teoría y Método*, 3, 85-127. Disponible en: https://doi.org/10.37417/RPD/vol_3_2021_535.

Soriano Arnanz, A. (2021c). La aplicación del marco jurídico europeo en materia de igualdad y no discriminación al uso de aplicaciones de Inteligencia Artificial. En P. R. Bonorino Ramírez, R. Fernández Acevedo y P. Valcárcel Fernandez (dirs.). *Nuevas normatividades: inteligencia artificial, derecho y género* (pp. 63-88). Cizur Menor: Aranzadi.

Soriano Arnanz, A. (2022). Discriminación algorítmica: garantías y protección jurídica. En L. Cotino Hueso (dir.) y M. Bauza Reilly (coord.). *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas* (pp. 139-170). Cizur Menor: Aranzadi.

Tobler, C. (2008). *Limits and potential of the concept of indirect discrimination*. European Network of Legal Experts in Anti-Discrimination. Disponible en: https://is.gd/aP67O1.

Tobler, C. (2022). *Indirect discrimination under Directives 2000/43 and 2000/78*. European Network of Legal Experts in Gender Equality and Non-Discrimination. Disponible en: https://is.gd/TA5LBc.

Xenidis, R. (2020). Turning EU equality law to algorithmic discrimination: Three pathways to resilience. *Maastricht Journal of European and Comparative Law*, 27 (6) 736-758. Disponible en: https://doi.org/10.1177/1023263X20982173.

Xenidis, R. y Senden, L. (2020). EU Non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination. En U. Bernitz, X. Grossout, J. Paju y S. De Vries (éds.). *General Principles of EU Law and the EU Digital Order* (pp. 151-182). Alphen aan den Rijn: Kluwer Law International.