Discriminació algorítmica. Límits de la regulació antidiscriminació europea i recomanacions preliminars

Anna Capellà Ricart
Universitat Autònoma de Barcelona
anna.capella@autonoma.cat
ORCID: 0000-0002-9898-8242

Recepció: març 2022
Acceptació: maig 2022
Publicació: juliol 2022

Cita recomanada: CAPELLÀ RICART, A. (2022). "Discriminació algorítmica. Límits de la regulació antidiscriminació europea i recomanacions preliminars". *Quaderns IEE: Revista de l'Institut d'Estudis Europeus*, *1*(2), 31-57.
DOI: <https://doi.org/10.5565/rev/10.5565/rev/quadernsiee.26>

Resum

Davant les discriminacions arran de les decisions automatitzades preses mitjançant sistemes algorítmics, la normativa antidiscriminatòria europea i la jurisprudència establerta fins ara, no semblen generar per si mateixes els mecanismes adequats per combatre-les. En aquest article repassem els tres eixos que presenten les principals limitacions per poder abastar aquest nou repte: una protecció excessivament basada en els atributs protegits; massa centrada en la dualitat discriminació directa o indirecta (que amb les noves eines causants de la discriminació poden veure's difuminades); i les dificultats a nivell probatori que pot tenir algú que ha patit discriminació algorítmica, com per exemple, trobar un grup amb qui comparar-se, o alguna persona o entitat a qui responsabilitzar de l'acció discriminatòria. Cal identificar les mancances i les formes de discriminació que escapen a la regulació actual per poder proposar mecanismes que realment serveixin en la lluita envers la discriminació.

Paraules clau: Unió Europea; Intel·ligència artificial; Algoritmes; Discriminació indirecta; Interseccionalitat; Igualtat de gènere.

Resumen. *Discriminación algorítmica. Límites de la regulación antidiscriminación europea y recomendaciones preliminares*

Ante las discriminaciones a raíz de las decisiones automatizadas tomadas mediante sistemas algorítmicos, la normativa antidiscriminatoria europea y la jurisprudencia establecida hasta ahora, no parecen generar por sí mismas los mecanismos adecuados

para combatirlas. En este artículo repasamos los tres ejes que presentan las principales limitaciones para abarcar este nuevo reto: una protección excesivamente basada en los atributos protegidos; demasiado centrada en la dualidad discriminación directa o indirecta (que con las nuevas herramientas causantes de la discriminación pueden verse difuminadas); y las dificultades a nivel probatorio que puede tener alguien que ha sufrido discriminación algorítmica, como por ejemplo, encontrar un grupo con el que compararse, o alguna persona o entidad a la que responsabilizar de la acción discriminatoria. Hay que identificar las carencias y formas de discriminación que escapan a la regulación actual para poder proponer mecanismos que realmente sirvan en la lucha hacia la discriminación.

Palabras clave: Unión Europea; Inteligencia artificial; Algoritmos; Discriminación indirecta; Interseccionalidad; Igualdad de género.

Abstract. *Algorithmic discrimination. The limits of EU antidiscrimination law and preliminary recommendations*

European anti-discrimination legislation and the jurisprudence established so far, do not seem to generate the appropriate mechanisms to combat discrimination as a result of automated decisions made through algorithmic systems. In this article we review the three axes that present the main limitations to face this new challenge: a protection overly based on protected attributes; overly focused on the duality of direct or indirect discrimination (that with the new tools that cause discrimination can be blurred); and the difficulties that someone who has been discriminated against may have at the probationary level, such as finding a group with whom to compare themselves, or who is responsible for the discriminatory action. It is necessary to identify the shortcomings and forms of discrimination that escape the current regulation to be able to propose mechanisms that really serve to fight against discrimination.

Keywords: European Union; Artificial intelligence; Algorithms; Indirect discrimination; Interseccionality.

Sumari

1. Introducció

2. Adaptació d'una normativa antidiscriminació europea basada en atributs protegits a la discriminació algorítmica

3. Difícil encaix dels conceptes discriminació directa i discriminació indirecta en els sistemes algorítmics

4. Dificultats per provar la discriminació algorítmica

5. Conclusions

Bibliografia

1. INTRODUCCIÓ

El concepte "decisió automatitzada", es defineix en el document anomenat "*Directrices...*"[1], com a aquella basada únicament en mitjans tecnològics sense la participació de l'ésser humà. Aquest tipus de decisions poden ser preses mitjançant intel·ligència artificial que, tal com està definida a l'Enciclopèdia Britànica, s'entén com la capacitat d'un ordinador o un robot controlat per ordinador per realitzar tasques comunament associades amb éssers intel·ligents, aplicant-se freqüentment a desenvolupar sistemes dotats dels processos intel·lectuals característics dels humans, com la capacitat de raonar, descobrir significats, generalitzar o aprendre d'experiències passades[2].

Les decisions automatitzades, tal com les decisions no automatitzades, causen situacions discriminatòries. Aquestes són causades pels errors i biaixos en les bases de dades d'entrenament dels algoritmes, per la selecció i pes que el sistema atorga a les variables i a les dades de les quals disposa, per l'obtenció de dades personals mitjançant correlacions amb altres dades (a través de l'ús de dades en aparença neutres[3] que revelen atributs personals i sensibles); per la perpetuació d'estereotips en els quals es basen les estructures socials de discriminació (donat que moltes de les decisions incloses en entrenar un algoritme són el resultat dels perjudicis implícits de la persona que els entrena, que al seu torn pot ser un reflex dels estereotips culturals que prevalen

[1] Grupo de Trabajo del articulo 29. Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679, Adoptadas el 3 de octubre de 2017. Revisadas i adoptades el 6 de febrero de 2018 (WP251), p.8. Disponible a: https://ec.europa.eu/newsroom/article29/items/612053

[2] "Artificial Intelligence". Britannica. 2021. Disponible a: https://www.britannica.com/technology/artificial-intelligence

[3] En aquest sentit, a les dades neutres les anomenarem indicadors (en anglès s'utilitza el terme "proxy", que vol dir apoderat), enteses per part de les persones que programen els sistemes algorítmics com a elements que estan a prop i per tant poden sevir com a aproximació a altres elements als quals no tenen accés directe.

en el seu entorn); o per la intencionalitat discriminatòria de les persones que les empren.

La Unió Europea (d'ara en endavant UE) pretén lluitar contra la discriminació en la definició i execució de les seves polítiques i accions[4] i regular les situacions discriminatòries que es trobin en el seu àmbit d'aplicació, protegint les persones de situacions en les que s'evidencien tractaments diferents a persones en base a unes categories protegides que es troben enumerades en l'article 19.1 del Tractat de Funcionament de la Unió Europea[5] (d'ara en endavant TFUE). Les previsions del dret derivat —que principalment es troben en les directives antidiscriminació[6]— estableixen un marc de garanties enfront les discriminacions que tots els Estats membres han d'integrar en els seus marcs legislatius -i que no impedeixen un desenvolupament més ampli-. A Espanya, la transposició de les directives s'ha fet a través de la Llei 62/2003, de 30 de desembre, de mesures fiscals, administratives i de l'ordre social[7], i de la Llei Orgànica 3/2007, de 22 de març, per a la igualtat efectiva d'homes i dones[8], incorporant d'aquesta manera els elements i salvaguardes del dret antidiscriminatori desenvolupats i avançats per la UE (Eguíluz, 2020). En el mateix sentit, a Catalunya, l'any 2020 es va aprovar la Llei 19/2020, del 30 de desembre, d'igualtat de tracte i no discriminació[9].

La problemàtica que se'ns presenta resideix en els riscos afegits per la discriminació que trobem en les decisions automatitzades i les mancances que

[4] Article 10 del TFUE. La discriminació contra la que pretén lluitar es aquella causada motius de sexe, origen racial o ètnic, religió o conviccions, discapacitat, edat o orientació sexual.

[5] L'article 19 del TFUE. Els motius que enumera aquest article son: el sexe, l'origen racial o ètnic, la religió o les conviccions, la discapacitat, l'edat o l'orientació sexual.

[6]Directiva 2000/43/CE relativa a l'aplicació del principi d'igualtat de tracte de les persones independentment del seu origen racial o ètnic; Directiva 2000/78/CE relativa a l'establiment d'un marc general per la igualtat de tracte en el treball i l'ocupació; Directiva 2006/54/CE relativa a l'aplicació del principi d'igualtat d'oportunitats i igualtat de tracte entre homes i dones en assumptes de treball i ocupació (refosa); Directiva 2004/113/CE per la qual s'aplica la igualtat de tracte entre homes i dones a l'accés de bens i serveis i al seu subministrament. Tant l'article 13 de la Directiva 2000/43/CE com l'article 8 de la Directiva 2004/113/CE, requereixen als Estats membres la designació d'organismes per assistir a les víctimes de discriminació, realitzar investigacions sobre les formes i prevalença de la discriminació i emetre informes i recomanacions. A España s'ha creat el Consejo para la Eliminación de la Discriminación Racial o Étnica (CEDRE) i L'Institut de la Dona, respectivament. Per una visió crítica respecte el CEDRE, vegeu: Rey Martínez, F. (2021). "O organismo español para a igualdade: claves para a súa converxencia co modelo europeo". *Revista Galega de Administración Pública (REGAP).* 43-73.

[7] Ley 62/2003, de 30 de diciembre, de medidas fiscales, administrativas y del orden social. Boletín Oficial del Estado, núm. 313, de 31 de diciembre de 2003.

[8] Ley Orgánica 3/2007, de 22 de marzo, para la igualdad efectiva de mujeres y hombres. Boletín Oficial del Estado, núm. 71 de 23 de marzo de 2007.

[9] Llei 19/2020, de 30 de desembre, d'igualtat de tracte i no-discriminació. Diari Oficial de la Generalitat de Catalunya, núm. 8307, de 31 de desembre de 2020. Aquesta llei tracta transversalment tot tipus de discriminacions (inclou a destacar com a característiques protegides –art. 1.3.h)- la condició social i econòmica, situació administrativa, professió o condició de privació de llibertat, i qualsevol manifestació d'aporofòbia o d'odi a les persones sense llar), deixant oberta la llista d'atributs concreta amb la següent fórmula: "Qualsevol altra característica, circumstància o manifestació de la condició humana, real o atribuïda, que siguin reconegudes pels instruments de dret internacionals" (art. 1.3.l). Es divideix en una part de foment que tracta les principals accions per evitar la discriminació en el si de l'administració i la societat i una part sancionadora.

s'observen en la normativa i jurisprudència europea actual per fer front a aquest nou repte. Volem abordar la temàtica basant-nos en tres eixos principals: una normativa antidiscriminatòria excessivament basada en els atributs a protegir (Secció 2), massa centrada en la dualitat entre discriminació directa i discriminació indirecta (Secció 3), i amb dificultats afegides a nivell probatori per evidenciar una discriminació algorítmica (Secció 4). Finalment, hi ha unes conclusions (Secció 5). L'objectiu de l'article és fer una anàlisi dels límits que presenta la normativa antidiscriminació europea per fer front a les discriminacions algorítmiques i, arran d'aquests límits, avançar algunes línies de possibles recomanacions per mitigar-los.

2. ADAPTACIÓ D'UNA NORMATIVA ANTIDISCRIMINACIÓ EUROPEA BASADA EN ATRIBUTS PROTEGITS A LA DISCRIMINACIÓ ALGORÍTMICA

En aquest apartat es volen defensar les següents idees: en primer lloc, precisar quins són els atributs que la normativa antidiscriminatòria europea protegeix fa que molts d'altres que es poden inferir a través de sistema de presa de decisió mitjançant algoritmes puguin quedar desprotegits (2.1). En segon lloc, que el fet que les directives antidiscriminació estableixin règims de protecció diferents, segons l'atribut protegit i l'àmbit sectorial d'aplicació, disminueix l'efecte de la protecció antidiscriminatòria, donat que molts sectors on l'anàlisi massiu de dades s'està duent a terme queden fora de l'àmbit d'aplicació —posant l'exemple de l'accés als béns i serveis— (2.2); i, en tercer lloc, atesa la situació, la necessitat imperant d'una perspectiva interseccional de la discriminació que no s'ha tingut en compte fins ara (2.3).

2.1. L'article 19 del TFUE i la discriminació algorítmica

En les decisions automatitzades mitjançant sistemes algorítmics tenen molta importància l'existència d'indicadors i correlacions estadístiques. En basar els resultats en una multitud de factors diferents i variables estadísticament relacionades, augmentarà els riscs de discriminació múltiple (que el TJUE encara no ha reconegut), és a dir, aquella basada en més d'un atribut protegit.

No només serà estrany que l'algoritme discrimini únicament en base a un motiu protegit sinó que ho podrà fer en base a atributs no estrictament protegits. Per exemple, si el sistema d'intel·ligència artificial crea noves categories que no es relacionen amb les característiques protegides concretes com l'estatus econòmic, la salut, l'educació o els ingressos (i altres factors que poden causar discriminació ja que es poden traduir en dades però no estan protegits), però que igualment diferencien les persones, essent aquestes injustes si reforcen les desigualtats socials (Zuiderveen, 2018); o si les categories protegides concretes no són suficientment representatives (per exemple, la categorització basada en el sexe pot ocasionar l'exclusió de persones que s'identifiquen

com a no binàries, transsexuals o intersexuals, forçant-los a enquadrar-se en categories amb els quals no s'identifiquen (Gerards i Xenidis, 20217).

En aquest sentit, s'ha fet palès que la redacció de l'article 19 del TFUE[10], que estableix una llista exhaustiva d'atributs protegits (sexe, origen racial o ètnic, religió o les conviccions, discapacitat, edat o orientació sexual) sense que sigui possible una ampliació per analogia a altres motius no expressament enumerats (Requena, 2011), potser té limitacions per abordar la discriminació algorítmica, i s'ha posat atenció en una redacció més oberta com aquella que es presenta en l'article 21 de la Carta de Drets Fonamentals de la Unió Europea (d'ara en endavant, Carta)[11] on es prohibeix tota discriminació (Gerards i Xenidis, 2021, p. 9).

2.2. Diferències rellevants en la protecció segons l'atribut protegit

Les directives antidiscriminatòries europees, que deriven del dret primari, estableixen protecció en base a la discriminació per raó de sexe, origen ètnic o raça, religió o creença, diversitat funcional, orientació sexual o edat[12].

Els límits en l'aplicació de les directives, ja sigui perquè es refereixen a atributs protegits determinats (i no a totes les formes de discriminació); ja sigui perquè es veuen limitades pel sector en el que s'apliquen (majoritàriament en relacions laborals i no en altres àmbits); o ja sigui perquè proporcionen excepcions que suposen un gran buit en relació a les decisions preses mitjançant algoritmes (com l'àmbit de la publicitat), són límits que ja suposaven llacunes en l'efectivitat de la legislació antidiscriminatòria, però que arran de la introducció dels sistemes algorítmics esdevenen més rellevants.

Agafem l'exemple de la protecció d'atributs protegits en l'accés als béns i serveis, ja que és un àmbit on s'empren sistemes de presa de decisió algorítmics. Només el sexe i —fins a un cert punt— l'origen ètnic, són atributs protegits més enllà del mercat (per exemple, seguretat social i sanitat en relació al sexe i educació en relació a l'origen ètnic). Dins del mercat estan també protegits més àmpliament que els altres atributs (la diversitat funcional, edat, orientació sexual, religió o creença), ja que estan protegits en l'àmbit de l'accés als béns i serveis en comptes de només en el context laboral. Això suposa que l'origen ètnic i el sexe estan protegits horitzontalment, és a dir, no només en les relacions público-privades que involucren l'Estat, sinó també en la relació entre

[10] L'article 19 del TFUE, en el seu primer apartat, disposa que sense perjudici d'altres disposicions dels Tractats i dins els límits de les competències atribuïdes a la Unió per aquests mateixos, el Consell, per unanimitat en base al procediment legislatiu especial i prèvia aprovació del Parlament Europeu podrà adoptar accions adequades per lluitar contra la discriminació per motiu de sexe, origen racial o ètnic, religió o creences, diversitat funcional o orientació sexual.

[11] L'article 21 de la Carta disposa que es prohibeix tota discriminació i, en particular, l'exercida per raó de sexe, raça, color, orígens ètnics o socials, característiques genètiques, llengua, religió o conviccions, opinions polítiques o de qualsevol altra mena, pertinença a una minoria nacional, patrimoni, naixement, discapacitat, edat o orientació sexual.

[12] Directiva 2000/43/CE; Directiva 2000/78/CE; Directiva 2006/54/CE; Directiva 2004/113/CE.

actors privats en el domini públic (activitats comercials i professionals), i en canvi, els altres atributs només es protegeixen en el context laboral (Xenidis, 2017).

En l'àrea de béns i serveis, la manca de protecció enfront la discriminació basada en edat, diversitat funcional, orientació sexual i religió o creences per part de la legislació antidiscriminatòria europea, representa un buit en un moment en que els algoritmes s'estan emprant per fer perfils de persones amb finalitats de venda (Gerards i Xenidis, 2021), multiplicant les oportunitats que les companyies privades facin perfils de clientela potencial mitjançant ofertes personalitzades. La discriminació algorítmica és, llavors, molt més procliu a ocórrer en el mercat de béns i serveis, ja que el seu propòsit és analitzar el comportament de les persones usuàries per donar una resposta a cadascuna d'elles amb diferents ofertes i oportunitats individualitzades, de manera que duent-se a terme la discriminació, grups i persones pertanyents a col·lectius desavantatjats poden ser exclosos legalment de l'accés a certs béns i serveis, tan importants com l'habitatge, la salut o l'educació (Xenidis, 2020).

A més, quant al sexe i la raça o origen ètnic, que es troben protegits horitzontalment, podem trobar certes disposicions que també dificulten la protecció. En referència al sexe, la Directiva 2004/113/EC exclou de la seva aplicació els mitjans de comunicació, la publicitat i l'educació. En vista de l'increment d'ús d'intel·ligència artificial en aquests camps, aquestes excepcions poden portar a debilitats en relació a la capacitat de la legislació europea a fer front a la discriminació algorítmica (Gerards i Xenidis, 2021). Quant a la raça o l'origen ètnic, en la Directiva 2000/43/EC no es fa cap excepció explícita com en la Directiva 2004/113/EC, però no es mencionen ni els mitjans de comunicació ni la publicitat. Seria lògic assumir que aquests camps estan coberts per la Directiva 2000/43/EC, però tot i així es considera interessant per part de Gerards i Xenidis (2021), donada la rellevància dels riscs de la discriminació algorítmica, la seva inclusió en la Directiva.

Malgrat que es podria legislar de manera més proteccionista a nivell nacional, i per tant, oferir més possibilitats de prevenir i redreçar la discriminació algorítmica, no hi ha cap prohibició harmonitzada a nivell europeu (Gerards i Xenidis, 2021). L'harmonització, o uns criteris estandarditzats i unes bases comunes —malgrat que es doni llibertat en l'aplicació—, seria especialment rellevant quan es fa front a uns sistemes que no tenen límits territorials i són essencialment globals, on la disparitat de legislacions i perspectives reguladores, segons el territori, poden disminuir la protecció.

2.3. Discriminació en base a múltiples criteris

La casuística de la discriminació algorítmica fa que la llista exhaustiva d'atributs de l'article 19 del TFUE i les exclusions en els àmbits d'aplicació de les directives antidiscriminació esdevinguin límits a la seva regulació. Un altre límit també molt vinculat a com es duu a terme la discriminació algorítmica és la manca de perspectiva interseccional.

D'una banda, s'entén interseccionalitat com la manera per la qual diferents tipus de discriminacions es relacionen o s'afecten entre si[13], i, d'altra banda, s'entén discriminació múltiple com qualsevol combinació de formes de discriminació contra persones per motiu de sexe, origen ètnic o racial, religió o creença, diversitat funcional, edat, orientació sexual, identitat de gènere o altres característiques, i també aquella patida per persones que tenen o es percep que tenen aquestes característiques[14]. Per tant, la interseccionalitat és una eina per a l'anàlisi de la discriminació, i la discriminació múltiple es produeix quan la persona és discriminada per més d'un motiu protegit.

Que no s'adopti una perspectiva basada en la interseccionalitat per abordar la discriminació fa més difícil la detecció i reconeixement de discriminació múltiple. La rellevància de reconèixer la discriminació múltiple quan ens referim a discriminació algorítmica té la seva base en l'ús d'indicadors i les correlacions estadístiques. Un algoritme rarament discriminarà de manera oberta i directa per un atribut protegit, ja que usualment basarà els seus resultats en una multitud de factors, les bases de dades seran diverses i serà difícil relacionar la discriminació amb un atribut protegit concret. És a dir, serà difícil identificar exactament quines variables expliquen un resultat algorítmic concret i més complicat encara detectar i aïllar el motiu de discriminació (Gillis i Spiess, 2019). La quantitat de dades creuades suposa un augment de la personalització dels resultats i, en conseqüència, un augment de les respostes especifiques donades a aquelles persones en les que concorren diverses categories que les fan especialment vulnerables (Soriano, 2021) —qüestió que s'abordarà en l'apartat 4.1.1.

A nivell jurisprudencial, malgrat la millor adaptació d'aquesta perspectiva per la lluita envers la discriminació algorítmica, el TJUE no ha mostrat predisposició a tenir-la en compte. En el cas *David L. Parris* contra *Trinity College Dublin*[15], el TJUE estableix que la disposició impugnada no és capaç de crear discriminació com a resultat de l'efecte combinat de l'orientació sexual i l'edat, ja que la norma no és discriminatòria ni en base a l'orientació sexual, ni en base a l'edat aïlladament[16]. De manera que

[13] "Intersectionality". Cambridge Dictionary. Cambridge University Press. 2021. Disponible a https://dictionary.cambridge.org/es/diccionario/ingles/intersectionality

[14] "Multiple discrimination". European Institute for Gendre Equality. 2021. https://eige.europa.eu/thesaurus/terms/1297

[15] TJUE. Sentència de 24 de novembre de 2016, *David L. Parris* v *Trinity College Dublin*, *Higher Education Authority, Department of Public Expenditure and Reform, Department of Education and Skills*, C443/15, EU:C:2016:897. La norma objecte d'estudi permet a les parelles unides civilment a rebre la pensió de viudetat si la unió s'havia format abans l'edat de 60 anys. El Sr. Parris no va poder contraure aquesta unió abans els 60 anys –malgrat estar en al relació durant 30 anys-, ja que les unions civils entre persones del mateix sexe no estaven permeses encara.

[16] En l'article 5.2 de la Llei 19/2020, de 30 de desembre, d'igualtat de tracte i no-discriminació s'especifica, en el mateix sentit que la sentència que "*En cas de discriminació múltiple, la justificació de la diferència de tracte s'ha d'acomplir respecte de cadascun dels motius de discriminació*". En relació a altres casos en que el TJUE no ha emprat la perspectiva interseccional vegeu: *Samira Achbita i Centrum voor gelijkheid van kansen en voor racismebestrijding* contra *G4S Secure Solutions NV*, de 14 de merç de 2017; i *Asma Bougnaoui* i *Association de défense des droits de l'homme (ADDH)* contra *Micropole SA*, de 14 de març de 2017.

s'analitzen les dues raons separadament per valorar si el Sr. Parris ha sofert un desavantatge particular (sentència que s'ha vist criticada en base al seu efecte perjudicial envers persones afectades per discriminació múltiple).

El TEDH, en canvi, en l'assumpte *Caravalho Pinto*, conclou que la qüestió que es planteja no és en relació amb l'edat o el sexe com a tals, sinó en la suposició que la sexualitat no és tan important per a una dona de cinquanta anys i mare de dos fills com per a una persona de menor edat, evidenciant la suposició que la sexualitat femenina està essencialment vinculada a les finalitats de procreació i sense tenir en compte la seva importància física i psicològica per a la realització personal de les dones com a persones[17]. Es pot comprovar en aquesta reflexió, com el TEDH aposta per analitzar la situació des de la perspectiva de les dues categories sensibles que es veuen en aquest cas implicades, com són l'edat i el sexe, i com el fet de tenir-les en compte juntes suposa que s'abordi el cas apreciant-ne millor la seva complexitat.

A nivell normatiu, en la Directiva 2000/43/EC i en la Directiva 2000/78/EC només alguns considerants es refereixen a la discriminació múltiple, més que en la Directiva 2004/113/EC i la Directiva 2006/54/EC, on no s'hi fa cap menció, de manera que actualment no es dona resposta a aquesta problemàtica i les situacions de discriminació (combinada) no es protegeixen. El Parlament Europeu en la seva "Resolució sobre la estratègia de la Unió per a la igualtat de gènere", de 21 de gener de 2021, admet que les polítiques de la Unió no han aplicat fins ara una perspectiva interseccional i s'han centrat principalment en la dimensió individual de la discriminació, sense abordar les seves dimensions institucionals, estructurals i històriques, afirmant que s'han d'abordar totes les formes de discriminació per aconseguir la igualtat de gènere per a totes les dones[18]. També en referència al gènere, en la comunicació de 5 de març de 2020, "Una Unión de la igualdad: Estrategia para la Igualdad de Género 2020-2025", de la Comissió Europea al Parlament Europeu, el Consell, el Comitè Econòmic i Social Europeu i el Comitè de les Regions, es fa èmfasi en que l'estratègia emprarà la interseccionalitat com a principi transversal per a treballar la igualtat de gènere[19].

En aquesta secció hem analitzat com és de limitant que no s'adopti una perspectiva interseccional en la discriminació algorítmica, ja que dificulta la possibilitat d'entendre com discrimina un sistema algorítmic (com té la capacitat, pot discriminar en base a molts atributs protegits interrelacionats), i, per tant, regular la discriminació que ocasiona adequadament. De la mateixa manera, també suposa un límit a la regulació una llista exhaustiva d'atributs protegits com la de l'article 19 TFUE i que en

[17] TEDH. Sentència de 25 de juliol de 2017. *Carvalho Pinto* contra Portugal. Núm. demanda 17484/15. Para. 52.

[18] Resolución del Parlamento Europeo, de 21 de enero de 2021, sobre la estrategia de la Unión para la igualdad de género (2019/2169(INI)). https://www.europarl.europa.eu/doceo/document/TA-9-2021-0025_ES.html

[19] Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Una Unión de la igualdad: Estrategia para la Igualdad de Género 2020-2025. COM/2020/152 final.
https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:52020DC0152

les directives antidiscriminatòries no s'estengui als mateixos àmbits la protecció de tots els atributs protegits. Seguidament, abordarem com les conceptualitzacions de discriminació directa i indirecta són de difícil adaptació a la discriminació algorítmica.

3. DIFÍCIL ENCAIX DELS CONCEPTES DISCRIMINACIÓ DIRECTA I DISCRIMINACIÓ INDIRECTA EN ELS SISTEMES ALGORÍTMICS

Leese (2014) planteja que les regulacions antidiscriminació només despleguen la plenitud del seu poder regulador quan s'apliquen a modes tradicionals de discriminació estàtics i no a aquells provocats per l'anàlisi automatitzat de dades, dinàmic i implícit, que escapa constantment de la regulació i potser ja no es pot abordar mitjançant la perspectiva de la discriminació directa i indirecta establerta fins ara. Es considera discriminació directa aquella situació en la qual una persona és tractada de manera menys favorable del que ha sigut o vagi a ser tractada una altra persona en una situació comparable a causa d'una categoria protegida —per exemple l'ètnia o el sexe— (Rey, 2019). En canvi, la discriminació indirecta —concepte creat pel Tribunal Suprem Federal dels Estats Units en la sentència Griggs contra Duke Power Company, de 8 de març de 1971[20]—, es podria definir com aquella situació en la que una pràctica que aparentment és neutra, acaba discriminant a persones en base a una característica protegida[21].

Primer abordarem perquè els dos conceptes per separat poden tenir un abast insuficient a l'hora d'abordar la discriminació algorítmica (3.1) i, seguidament, plantejarem, com inclús pot ser que es difumini la frontera que s'ha establert entre ells (3.2), donada la metodologia discriminatòria dels sistemes algorítmics.

3.1. Problemàtiques de la discriminació algorítmica directa i discriminació algorítmica indirecta

La discriminació directa la podem trobar quan es veuen implicats sistemes algorítmics, per exemple, quan aquests decideixen que les dones no són adequades per a un lloc de treball, ja que l'algoritme ha estat entrenat amb bases de dades on no hi ha pràcticament currículums de dones. Per tant, l'algoritme atorgarà al fet de ser dona un valor més negatiu, en comparació al fet de ser home a causa de la seva infrarrepresentació en la base de dades, i arribarà a la conclusió que es preferible contractar homes que dones. En conseqüència, el valor negatiu associat a la categoria de dades podrà determinar de manera directa el resultat (Soriano, 2021).

[20] Tribunal Suprem dels Estats Units. Sentència de 8 de març de 1971. *Griggs* v. *Duke Power Co.*, 401 U.S. 424.

[21] Article 2(2)(b) de la Directiva 2000/43/CE; Article 2(2)(b) de la Directiva 2000/78/CE; Article 2(1)(b) de la Directiva 2006/54/CE (refosa); i Article 2(b) de la Directiva 2004/113/CE.

La discriminació indirecta —que té com a requisits l'existència d'un criteri, disposició o pràctica aparentment neutre, que creï un desavantatge particular d'un grup respecte d'un altre, i que no sigui justificable objectivament amb una finalitat legitima— la podem trobar quan es veuen implicats sistemes algorítmics, per exemple, per predir quins deutors tindran problemes per pagar els préstecs (objectiu final fixat). El sistema, que no té informació sobre l'ètnia, aprèn que les persones que pertanyen a cert codi postal tendeixen a tenir més problemes per pagar els préstecs (factor diferencial entre aquelles persones que paguen el préstec i aquelles que no) i ho utilitza com a criteri de predicció per a futures persones susceptibles d'impagaments (fixant de manera autònoma aquest criteri). Tot i semblar un criteri neutre, pot existir una relació directa entre el codi postal i l'ètnia, que suposi que el banc denegui préstecs a persones amb aquest codi postal, perjudicant, a la pràctica, persones amb un origen ètnic determinat (Zuiderveen, 2018).

Hi ha dificultats d'adaptació a la normativa antidiscriminatòria que es poden trobar en ambdós casos. L'ús de variables i categorització de dades en els algoritmes d'aprenentatge automàtic està en constant evolució mentre l'algoritme aprèn. Com aquestes categories no són estàtiques, serà difícil saber si estan relacionades amb els atributs protegits. S'haurà de mirar el model algorítmic i la manera com el model estadístic usat tracta les dades disponibles durant el temps per trobar com el tractament desfavorable actua, una situació que pot ser difícil d'acomplir donats els problemes d'accessibilitat (Gerards i Xenidis, 2021). Aquí entren en conflicte la categorització d'atributs protegits estàtica que ofereix la llei (ja esmentada en l'apartat 2.1) i la perspectiva dinàmica i matemàtica dels algoritmes per a la categorització de dades. També, l'opacitat del sistema dificulta l'accés a proves de l'existència de discriminació (o inclús la consciència que s'està patint discriminació —problema principal de la discriminació algorítmica—) i es poden magnificar i diversificar els efectes discriminatoris en tenir la possibilitat d'incloure en la decisió dades d'una gran quantitat de persones.

Concretament, en relació amb la discriminació algorítmica directa, s'ha discutit que els atributs categoritzats com a protegits usualment no s'empraran directament en els algoritmes. Això pot passar, o bé, per evitar la detecció de discriminacions, donat que és molt més fàcil eludir la regulació antidiscriminació fent ús d'indicadors que representin atributs protegits (provar que l'ús dels indicadors és intencionat és molt difícil i els efectes seran idèntics a la discriminació no intencional que pot resultar de l'anàlisi de dades (Páez, 2021)); o bé, per evitar una disminució de la precisió algorítmica, que s'incrementa com més correlacions diverses es poden fer (Hacker, 2018). Per tant, la discriminació directa podria ser menys probable en la discriminació algorítmica que en la discriminació duta a terme per éssers humans (Gerards i Xenidis, 2021). També, el tractament de les dades i la seva categorització, per part dels algoritmes, pot no ser cognoscible per a l'individu i, per tant, ser impossible determinar on hi ha la discriminació directa (Leese, 2014). Addicionalment, trobem l'element dissuasiu que ens fa pensar que la discriminació directa disminuirà quan s'emprin

algoritmes: s'ha posat l'atenció en el fet que a mesura que les obligacions legals augmenten, la discriminació directa es procliu a disminuir en la fase de desenvolupament dels algoritmes (Hacker, 2018).

En el cas de la discriminació algorítmica indirecta, probablement esdevindrà difícil determinar exactament quin és el criteri, disposició o pràctica neutra que causa la discriminació, ja que sense accés al sistema per part de la persona que s'ha considerat discriminada, no es pot determinar quina és la variable (o variables) que podria causaria l'efecte discriminatori. També cal considerar que com la discriminació indirecta es centra en el resultat més que en el tracte, és necessari evidenciar l'impacte de la mesura que es posa en qüestió entre dos grups, el de la presumpta víctima i un altre. Aquest exercici de comparació, quan estan implicats sistemes algorítmics, pot resultar molt complicat (s'abordarà aquesta qüestió en l'apartat 4.3.3).

Endemés, les justificacions objectives en casos de discriminació indirecta no són *numerus clausus*, contràriament als casos de discriminació directa[22]. La justificació de la discriminació indirecta es basa en provar que la mesura discriminatòria persegueix una finalitat legítima, permetent al tribunal realitzar un examen de proporcionalitat a través del qual es ponderin els interessos en joc. En general, el TJUE[23] ha donat per vàlides pràcticament totes aquelles finalitats que no suposin un resultat discriminatori o que amaguin l'objectiu de discriminar de manera evident (Tobler, 2008).

Tenint en compte l'opacitat i complexitat de la presa de decisions mitjançant sistemes algorítmics, és poc probable que es pugui analitzar si totes les variables que es tenen en compte són necessàries per a la presa de la decisió o si prescindint d'elles també es podria arribar al mateix resultat; si la base de dades emprada és suficientment representativa o conté biaixos; o si el pes que se li dona a cada variable és adequat, resultant més probable que el test d'adequació es faci en relació al sistema en el seu conjunt i no separadament (Hacker, 2018).

L'adequació d'un sistema de presa de decisió amb algoritmes com a mesura per aconseguir una finalitat establerta segurament dependrà de l'existència d'un altre algoritme que pugui aconseguir el mateix sense tenir uns efectes tan discriminatoris, ja que actualment no existeixen altres maneres de processar la quantitat de dades que tracten els algoritmes en la mateixa quantitat de temps. La part demandada, en teoria, haurà d'acreditar que va valorar l'ús d'altres sistemes i bases de dades i que no existia un algoritme menys discriminatori, ni una base de dades més completa i menys

[22] Cal puntualitzar aquí que a diferència de la discriminació directa on el TJUE estableix justificacions taxades i, en canvi, el TEDH, té un sistema de justificació general objectiva; en el cas de la discriminació indirecta, tant el TJUE com el TEDH estableixen un sistema de justificació general objectiva.

[23] Vegeu: TJUE (Gran Sala). Sentencia de 13 de març de 2017, *Samira Achbita i Centrum voor gelijkheid van kansen en voor racismebestrijding* contra *G4S Secure Solutions* NV, C-157/15, EU:C:2017:203. Para. 38; TJUE. Sentència de 13 de maig de 1986, *Bilka - Kaufhaus GmbH* v *Karin Weber von Hartz*, C-170/84, EU:C:1986:204. Para. 37 i 38. Ens referim a "pràcticament totes les finalitats" ja que hi ha certs casos particulars, com per exemple en relació a les finalitats llegítimes en els sector públic, en que el TJUE ha determinat que les consideracions de caràcter pressupostari no poden en si mateixes suposar una finalitat llegítima que justifiqui una mesura indirectament discriminatòria. Vegeu: TJUE (Gran Sala). Sentència d'11 de novembre de 2014, *Leopold Schmitzer* contra *Bundesministerin für Inneres*, C-530/13, EU: C: 2014:2359, para. 41.

esbiaixada. Tot i així, s'haurà de tenir en compte també la capacitat tècnica i econòmica de la part demandada (Soriano, 2021), establint-se d'aquesta manera un ampli marge d'apreciació.

En aquest apartat hem analitzat quins límits trobem tant en la discriminació directa com en la discriminació indirecta per regular la discriminació que es produeix mitjançant els sistemes algorítmics. Seguidament, analitzem com la discriminació que es produeix mitjançant sistemes algorítmics pot acabar difuminant la frontera que existeix entre els dos conceptes, en el cas que es produeixi la superposició del significat de certs indicadors amb aquell dels atributs protegits.

3.2. Pot difuminar-se la frontera entre discriminació directa i indirecta?

Com la discriminació algorítmica actua de manera més subreptícia, donada la opacitat dels sistemes algorítmics, se la pot enquadrar millor en la conceptualització de discriminació indirecta que en la conceptualització de discriminació directa. En primer lloc, a través dels sistemes algorítmics es perpetuen accions discriminatòries no intencionals[24] i s'incrementa el seu impacte (Barocas, 2016). En segon lloc, malgrat l'existència de proteccions legals més exigents per les categories especials de dades, els sistemes d'intel·ligència artificial tenen gran capacitat per crear noves vies alternatives menys evidents per inferir atributs protegits d'altres dades, fent que els casos de discriminació directa s'augurin estranys en els sistemes automatitzats, ocasionant, en canvi, el risc d'una discriminació indirecta generalitzada i subtil (Wachter, Mittelstadt i Russell, 2021). I finalment, en tercer lloc, gràcies a la complexitat i opacitat de la interrelació de dades d'aquests sistemes, és molt més fàcil amagar la voluntat de dur a terme accions discriminatòries per part d'aquelles persones que prenen les decisions, ja que poden al·legar l'existència d'un algoritme defectuós com a manera d'eludir la seva responsabilitat.

Aquests arguments s'han de matisar, ja que la discriminació algorítmica pot difuminar la línia entre discriminació directa i indirecta portant a un repte de classificació, donat que serà cada cop més difícil identificar tractament diferencial basat en atributs protegits en el context de les operacions algorítmiques (Gerards i Xenidis, 2021). En aquest sentit, conclouen Barocas i Selbst (2016) que la discriminació directa i la discriminació indirecta esdevenen essencialment el mateix des d'una perspectiva probatòria. La qüestió resideix en si s'aborden els motius protegits des d'una perspectiva restrictiva o àmplia.

Més enllà de la identificació de característiques protegides en les decisions preses mitjançant algoritmes, existeixen les correlacions i els indicadors que tenen

[24] Biaixos implícits i inconscients que es donen quan alguna persona ha internalitzat un estereotip social de tal manera que aplica a una persona un tractament ostensiblement més desfavorable que a una altra. Les persones els hem anat adoptant a causa de viure en una societat desigual i marcada per les discriminacions sistematitzades i estructurals però de les quals molt probablement no en som conscients.

rellevància en relació a la definició mateixa d'atribut protegit i qüestionen com d'estretament o àmpliament s'haurien d'entendre. Si se'ls dona una interpretació extensiva, els indicadors més rellevants també haurien d'estar coberts per la protecció de discriminació directa, ja que, malgrat no ser exactament l'atribut, es produeix una superposició del significat quasi completa. L'exemple clàssic és la discriminació basada en l'embaràs, que està clarament relacionada amb la discriminació basada en el sexe, ja que embaràs es un indicador de "dona". El TJUE ha tractat la discriminació per embaràs com a discriminació directa[25]. Un altre bon exemple és el passaport estranger com a indicador de tenir una altra nacionalitat.

Usualment hi ha una clara superposició entre el motiu protegit i els seus indicadors, és a dir, que l'ús de l'indicador cobreix quasi exactament el mateix grup de persones que cobriria el motiu protegit. Contràriament, si s'entén l'atribut protegit de manera restrictiva, només quan es discrimini en base a aquest estarem parlant de discriminació directa i tota la discriminació que provingui dels indicadors o correlacions sobre aquest atribut es considerarà discriminació indirecta. Per tant, un repte important que sorgeix amb la discriminació algorítmica és descobrir quins indicadors arriben a tenir un grau tan gran de superposició amb el motiu protegit que poden ser vistos com la mateixa cosa. S'ha de qüestionar, llavors, si és necessari un 100 % de superposició o estadísticament també s'admet un indicador que tingui un 90 % o 80 % de superposició amb un atribut protegit i si aquest tractament desigual pot seguir considerant-se discriminació directa o s'ha de considerar indirecta o discriminació per associació[26] amb el grup que es caracteritza amb l'atribut protegit (Gerards i Xenidis, 2021).

També cal tenir en compte l'opacitat inherent als algoritmes. Quan hem d'abordar la caixa negra que suposen molts sistemes algorítmics, cercar tractament diferencial basat en atributs protegits pot esdevenir molt complicat i pot ocasionar que, en no localitzar si la discriminació prové directament d'un atribut sensible (sexe, origen ètnic, etc.) o d'un indicador (embaràs, passaport, etc.), s'acabi sense poder determinar exactament si es tracta de discriminació directa o indirecta, fent que al final sigui la discriminació indirecta la que actuï com a refugi per captar la discriminació algorítmica. Això pot suposar la reducció de la seguretat jurídica, si significa que es podrà aplicar el sistema de justificacions objectives obertes en comptes del conjunt *numerus clausus* de justificacions disponibles per la discriminació directa (Gerards i Xenidis, 2021).

En aquesta secció hem exposat els límits que podem trobar en els conceptes, tant de discriminació directa com de discriminació indirecta, per enquadrar la majoria de casos de discriminació algorítmica i com, inclús, la funcionalitat específica de la discriminació algorítmica ocasiona que es difumini la frontera que existeix entre ells.

[25] Vegeu: TJUE. Sentència de 8 de novembre de 1990, *Elisabeth Johanna Pacifica Dekker* contra *Stichting Vormingscentrum voor Jong Volwassenen (VJV-Centrum) Plus*, C-177/88, EU:C:1990:383.

[26] Aquells supòsits en el que l'algoritme no seleccioni a les persones per pertànyer a un grup desavantatjat, sinó per la seva relació amb persones que pertanyen a un grup desavantatjat.

Analitzarem, tot seguit, quins elements de prova de discriminació són de difícil adaptació a la discriminació algorítmica i per què.

4. DIFICULTATS PER PROVAR LA DISCRIMINACIÓ ALGORÍTMICA

Emprar procediments judicials per fer complir els principis d'igualtat i no discriminació en els algoritmes està ple de dificultats. No només per a les persones expertes pot ser difícil identificar la font de la discriminació algorítmica, sinó que qualsevol evidència de discriminació en el procés pot haver desaparegut quan l'algoritme ha "evolucionat". També, malgrat les persones expertes puguin localitzar la font de discriminació, pot ser impossible fer-ho per part de les persones que s'han vist discriminades, la judicatura, les persones legisladores, etc. (Gerards i Xenidis, 2021). La manca de casos que arriben als tribunals ens fa pensar que la complexitat de la legislació que habilitaria a les persones demandants per accedir-hi ho fa altament improbable, malgrat es produeixin les discriminacions. Cal llavors analitzar en què es concreta la complexitat probatòria per poder evidenciar les discriminacions algorítmiques. Hem dividit en tres apartats les problemàtiques: el repartiment de la càrrega de la prova (4.1), la manca de claredat en la responsabilitat de l'acció discriminatòria (4.2) i la pràctica absència de *tertium comparationis* (4.3).

4.1. Repartiment de la càrrega de la prova en els casos de discriminació algorítmica

Tant en el cas de la discriminació directa com indirecta, la jurisprudència determina que havent-se aportat indicis de discriminació *prima facie*, s'inverteix la càrrega de la prova i és la presumpta persona discriminadora qui ha de demostrar que no està realitzant cap conducta discriminatòria. La inversió de la càrrega de la prova pot ocórrer quan la persona demandant ha establert una discriminació *prima facie* o quan el nexe causal entre l'atribut protegit i el dany és probable (Ringelheim, 2019). També es pot invertir la càrrega de la prova si la part demandada es nega a oferir la informació rellevant necessària per tal que les persones demandants puguin construir un cas, com passa en el cas Danfoss, sentenciat pel TJUE[27]. Aquesta manca de transparència pot justificar que el tribunal canviï la càrrega de la prova, forçant la part demandada a respondre (Farkas i O'Farrel, 2015). Assumint que el canvi de la càrrega de la prova

[27] TJUE. Sentència de 17 d'octubre de 1989. *Handels-og Kontorfunktionærernes Forbund i Danmark* contra *Dansk Arbejdsgiverforening* en nom de *Danfoss*. C- 109/88. En aquesta sentència s'estableix que quan una empresa aplica un sistema retributiu caracteritzat per la falta total de transparència, recau sobre la persona empresària la càrrega de provar que la seva política de salaris no es discriminatòria (para. 16); quan les dones hagin demostrat a partir d'un número relativament important que la seva retribució mitja és inferior a la dels homes. En aquest cas, malgrat s'estableix que és l'empresa qui ha de determinar que el seu sistema no és discriminatori (es podria extrapolar als algoritmes —l'empresa ha de demostrar que l'algoritme no discrimina—), no es tracta d'una condició prèvia a la seva implantació, sinó que s'han d'aportar primer proves de que el sistema és discriminatori.

ocorre, en general hi ha dues maneres a través de les quals la persona demandada pot defensar-se: o bé, refutant que existeixi un nexe causal entre els resultats diferencials i l'atribut protegit[28]; o bé, reconeixent que els resultats diferencials han ocorregut, però es pot donar una justificació, en base a un interès legítim, de manera necessària i proporcionada[29].

Malgrat iniciar-se a nivell jurisprudencial, la doctrina de la necessitat d'indicis per invertir la càrrega de la prova de discriminació també es troba descrita, per exemple, en l'article 19.1 de la Directiva 2006/54, on s'especifica que les persones que han patit un perjudici causat per la no aplicació del principi d'igualtat de tractament, hauran de presentar davant l'autoritat competent fets en relació als quals es pot presumir que ha existit una discriminació i, llavors, la persona demandada haurà de demostrar que no s'ha incomplert el principi d'igualtat de tracte (Kullmann, 2018).

En l'assumpte *Galina Meister* contra *Speech Design Carrier Systems GmbH*[30], el TJUE va determinar que la negativa de l'empresa a donar informació —que impedia *per se* que la persona pogués establir un cas *prima facie* de discriminació— podia comptar per l'establiment del cas de *prima facie*. També va refusar l'existència d'un dret a la informació per part de la persona que presumptament havia patit discriminació. Això significa que la negativa de donar informació pot establir un cas *prima facie*, però també, en canvi, que no està garantida la possibilitat de les víctimes —i inclús els òrgans de supervisió— de reunir les proves necessàries. Tant en la discriminació directa com indirecta, malgrat tenir la informació també és rellevant comprendre el contingut dels sistemes per plantejar la prova. En aquest sentit, el tribunal tindrà molt difícil accedir a informació rellevant per poder prendre una decisió si l'algoritme és opac. El TJUE, per exemple, en l'assumpte *Elisabeth Johanna Pacifica Dekker* contra *Stichting Vormingscentrum voor Jong Volwassenen (VJV-Centrum) Plus* —ja esmentat—, va determinar que rebutjar una única candidata a una posició de feina a causa d'estar embarassada era discriminació directa per raó de sexe, ja que només es pot donar el cas si la persona candidata és dona[31]. Aquesta situació és difícilment comparable en el cas que la decisió sigui presa per un algoritme al contingut del qual no es pugui accedir.

En l'assumpte *SyRI*, es demana a la part demandada que aporti informació sobre el model de risc i els indicadors dels quals es composa com també informació objectivament verificable sobre la naturalesa del sistema algorítmic. La part demandada deliberadament es nega a aportar aquesta informació i això ocasiona que

[28] També es pot emprar aquesta defensa en la discriminació directa.

[29] TJUE. Sentència de 13 de maig de 1986, *Bilka - Kaufhaus GmbH* v *Karin Weber von* Hartz, C-170/84, EU:C:1986:204. Para 45.

[30] TJUE (Segona Cambra). Sentència de 19 d'abril de 2012. *Galina Meister* contra *Speech Design Carrier Systems GmbH*. C-415/10. EU:C: 2012:217. En aquest cas, La persona demandant al·legava discriminació per raó de sexe, edat i origen ètnic en relació a un procés de selecció de personal i volia obtenir informació sobre quines altres persones candidates eren de l'empresa de selecció.

[31] TJUE. Sentència de 8 de novembre de 1990. *Elisabeth Johanna Pacifica Dekker* contra *Stichting Vormingscentrum voor Jong Volwassenen (VJV-Centrum) Plus*. C177/88. EU: C: 1990:383. Para 12.

el Tribunal del Districte de la Haia acabi considerant el sistema contrari a l'article 8.2 de la Convenció Europea de Drets Humans[32].

En el cas de la discriminació mitjançant sistemes algorítmics, donada la opacitat i l'asimetria d'informació relativa al funcionament del sistema, a una persona li pot ser impossible aportar indicis (en molts casos necessitarà informació grupal que li serà inaccessible, ja que no tindrà cap tipus de connexió amb la resta de persones que s'han vist discriminades de la mateixa manera no podent acreditar un nexe d'unió, o que es forma part d'un grup desavantatjat en base a una característica comuna). S'hauria de plantejar si en el cas de la discriminació algorítmica, donada l'asimetria de poder, s'hauria d'automatitzar una inversió de la càrrega de la prova quan la persona no pugui aportar informació -assimilant-ho a la negativa de l'empresa a donar informació-, ja que la persona que s'ha vist discriminada probablement es trobarà sense suficient informació (i si la té, sense eines comprendre-la) per establir un cas *prima facie*.

4.2. La responsabilitat de la discriminació algorítmica

Tradicionalment la responsabilitat de la discriminació la té la persona que duu a terme l'acció discriminatòria ja sigui directa o indirectament. En el cas dels sistemes algorítmics hi ha una multiplicitat de factors, per exemple, biaixos implícits en les persones que desenvolupen, vetllen pel funcionament i apliquen les decisions preses pel sistema algorítmics, bases de dades esbiaixades, una voluntat discriminatòria de l'organització que empra el sistema, múltiples actors, etc., que poden contribuir a dificultar l'establiment de la responsabilitat.

El que ocasiona la dificultat de trobar una persona responsable es pot concretar en que hi ha tantes persones i organitzacions involucrades en la planificació, el desenvolupament i l'ús dels algoritmes que pot no ser fàcil saber a qui adreçar una demanda davant un tribunal. Diferents persones o organitzacions *En la Secció 2, considerem* són responsables d'establir els objectius, construir el procés de decisió, programar i entrenar l'algoritme, recol·lectar i preparar les dades, emprar els algoritmes en la decisió o monitoritzar i supervisar el seu efecte[33]. Si la discriminació sorgeix al final de la cadena, com es pot traçar la responsabilitat? Els algoritmes s'empren habitualment per finalitats comercials i estan protegits pel dret a la propietat intel·lectual i industrial, podent minar les exigències de transparència i complicar l'exercici de recerca de responsabilitats. Actualment no hi ha cap obligació per a una companyia de donar accés als algoritmes o revelar les seves característiques, podent haver creat el sistema amb una base de dades accessible públicament, però de propietat

[32] Tribunal de Districte de La Haia. Sentencia C/09/550982/ HA ZA 18-388, de 5 de febrer de 2020, *Case of SyRI. F. J.* 6.49. https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878

[33] Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems, Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers' Deputies. para. 14.

privada pel que fa al valor de les dades i el model creat a partir de la base (Crawford, 2021, p. 120).

Els algoritmes d'aprenentatge automàtic o d'aprenentatge profund (i inclús a vegades els que es recolzen en normes que es basen en processos de decisió complexos) són caixes negres i els seus procediments no són intel·ligibles per a les persones, o inclús aquelles expertes en aquest àmbit, i es fa difícil traçar i aïllar la causa del resultat discriminatori. Per exemple, les guies de l'Agencia Espanyola de Protecció de Dades estableixen que una persona que pren decisions recolzant-se en la intel·ligència artificial no pot evitar la responsabilitat mitjançant l'argument que hi ha insuficient accés a la informació o coneixement tècnic en relació al funcionament del sistema[34].

Malgrat això, moltes vegades donada la opacitat subsisteixen dubtes difícils o impossibles de resoldre per a persones no especialitzades però inclús per persones especialitzades, per exemple, quina correlació és problemàtica? El problema ve de les dades esbiaixades o el disseny algorítmic reflecteix un estereotip perjudicial? Inclús si es pot localitzar la font del problema, els algoritmes d'aprenentatge automàtic estan en constant evolució i per tant, la font de discriminació pot canviar al llarg del temps (Gerards i Xenidis, 2021, p. 74). Com, en aquestes circumstàncies, es pot concretar la responsabilitat?

En el cas que l'empresa, organització o administració utilitzi el seu propi model ja serà difícil provar que una alternativa millor existeix, però esdevé un problema irresoluble quan els models emprats són de terceres parts, i la persona de qui depèn finalment la decisió no té cap possibilitat de retocar el model. En aquest cas, a la persona que ha pres la decisió únicament podrà reclamar-se-li responsabilitat per haver contractat una empresa que empra un model discriminatori. Aquest seria l'exemple de persones empresàries que utilitzen *Facebook* o *HireVue* per contractar, ja que en no utilitzar un model propi, només se les podrà fer responsables d'haver contractat un sistema de selecció –que s'haurà de provar, en el cas que no estigui provat- que és discriminatori, i per tant serà molt fàcil per la persona que ha pres la decisió eludir la seva responsabilitat (Gerards i Xenidis, 2021, p. 27).

4.3. Absència de tertium comparationis quan ens referim a les discriminacions algorítmiques

En aquest apartat abordarem els límits que genera la perspectiva individual de la protecció envers la discriminació —no adaptada a les discriminacions grupals algorítmiques— (4.3.1); com el grau de personalització que es dona en la presa de decisions mitjançant sistemes algorítmics ocasiona —específicament en la discriminació indirecta—, la pràctica inexistència d'un grup desavantatjat —trobant-

[34] Agencia Española de Protección de Datos (AEPD). (2020). Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. p. 97. https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf.

nos amb moltes persones perjudicades sense eines per interrelacionar-se entre elles—(4.3.2); i com, a més a més de les dues qüestions anteriors, la manca de mitjans per accedir a informació d'altres o informació relativa a grups de població genera a la pràctica la impossibilitat d'establir un grup comparador (4.3.3).

4.3.1. Protecció individual, litigació individual

La normativa antidiscriminatòria europea es basa en la litigació individual. Com ja hem dit, la persona que presumptament s'ha vist discriminada té la càrrega d'aportar un cas *prima facie*[35].

Per aportar el cas *prima facie* —sigui en el cas de discriminació directa o sigui indirecta—, es necessita informació, i actualment és problemàtica l'obtenció de dades que puguin concretar-se en una prova per part de les persones que s'han vist discriminades. L'accés al que es refereix el Reglament General de Protecció de Dades (articles 13, 14 i 15 RGPD) sempre és a nivell individual, ja que es protegeixen les dades personals d'altres persones. Aquell individu que s'ha vist discriminat no podrà accedir a les dades personals d'altres perquè estaran protegides per la legislació de protecció de dades (Rivas-Vallejo, 2021; Hacker, 2018). També s'ha debatut en la doctrina l'existència d'un dret a l'explicació dels algoritmes arran del RGPD, que al ser individual, també esdevé insuficient (Roig, 2020). Com expliquen Weatherall i Caetano (2020), en el context de la normativa antidiscriminació, explicar decisions individuals pot no ajudar a identificar la discriminació indirecta: informar sobre el funcionament del model en conjunt i la distribució de resultats també és important. Endemés, malgrat ser lògic pensar que es tindran moltes proves estadístiques, donat que s'analitzen dades massivament, si la persona perjudicada no les compren i/o no sap explotar-les, no podran ser útils a l'hora de provar la discriminació (Xenidis i Senden, 2020)

La perspectiva individual de la protecció no pot fer front als sistemes algorítmics que també tenen la capacitat d'excloure directament les persones. La personalització de la publicitat que ofereixen, pot ser directament contrària a la regulació antidiscriminatòria, ja que no dona accés a certes ofertes de feina segons les característiques que et defineixen. A *Feryn*, *Accept i Associazione Avvocatura per í diritti LGTBI*, el TJUE va determinar que dissuadir les persones candidates pertanyents a grups protegits de presentar-se a un lloc de treball es considerava discriminatori malgrat no existís un procés de selecció en curs. La falta o la reducció de publicitat a persones integrants de grups protegits podria considerar-se discriminació si, en efecte, soscava l'objectiu de la legislació europea antidiscriminatòria de garantir l'accés igual al mercat de treball (Gerards i Xenidis, 2021). Sembla necessària una perspectiva col·lectiva —actualment inexistent—, i unes exigències de transparència mínimes per

[35] Vegeu en aquest sentit: TJUE. Sentència de 10 de juliol de 2008, *Centrum voor gelijkheid van kansen en voor racismebestrijding* contra Firma *Feryn NV*, C-54/07, EU:C:2008:397; TJUE, 25 d'abril de 2013, *Asociația Accept* contra *Consiliul Național pentru Combaterea Discriminării*, C-81/12, EU:C:2013:275; TJUE, Sentència de 23 d'abril de 2020, *NH* contra *Associazione Avvocatura per i diritti LGBTI - Rete Lenford*, C 507/18, EU:C:2020:289.

poder fer front a aquesta tipologia de discriminació. Addicionalment, el funcionament de la discriminació algorítmica també dificulta formar un grup desavantatjat i un grup amb qui comparar-se.

4.3.2. Impediments per trobar un grup desavantatjat

La personalització de tracte a les persones subjectes de la decisió automatitzada, mitjançant el sistema algorítmic, serà decisiva en la dificultat de trobar persones amb qui crear un grup desavantatjat. Tal com la discriminació directa pot provar-se a nivell individual a través de la referència explicita a una característica protegida a la norma impugnada (malgrat moltes altres dificultats, no hi ha necessitat de trobar un grup amb qui identificar-se), la discriminació indirecta si que requereix comparació a nivell grupal (Wachter, Mittelstadt i Russell, 2021). Per consegüent, la problemàtica que expliquem seguidament es dona preferentment en la discriminació indirecta. El sistema algorítmic de l'Agència Estatal de Treball d'Àustria (AMS) té una forta tendència a individualitzar les causes de desocupació (Allhutter, 2020). Això suposa que com l'algoritme té accés a tantes dades pot crear respostes personals per a cada persona que analitza, de manera que no només serà difícil trobar altres amb qui comparar-se, sinó que també serà difícil trobar elements comuns que hagin dut a l'algoritme a prendre una decisió discriminatòria. Al no existir un criteri que s'aplica de la mateixa manera a totes les persones, no hi ha una base per establir que un tracte diferencial ha estat discriminatori i, en no haver-hi casos comparables, és gairebé impossible establir un grup perjudicat.

En la discriminació indirecta, esdevé molt complicat identificar-se dins d'un grup que s'ha vist desavantatjat, ja que pot ser que no existeixin dos casos en els que s'hagi actuat de manera idèntica. Ha de ser evident que un grup que comparteix una o diverses característiques s'ha vist en una posició desavantatjada en relació a la decisió presa i que les persones d'aquest grup puguin conèixer que existeixen altres en la mateixa situació. Els sistemes de presa de decisió mitjançant algoritmes habitualment discriminaran en base a múltiples factors i serà molt complicat erigir-se com a grup desavantatjat concret quan potser els motius de discriminació no són tots els mateixos i definir-se com a grup pot ser confús.

Endemés, normalment el grup desavantatjat està integrat per persones que estan protegides per un motiu de discriminació de manera predominant, però no exclusivament (Tobler, 2008). Per exemple, en l'assumpte del TJUE, *CHEZ Razpredelenie Bulgaria* contra *Komisia za zashtita ot diskriminasis*, de 16 de juliol de 2015[36], en un barri poblat majoritàriament per persones d'ètnia romaní —on tenia la demandant una botiga d'alimentació—, els cables de distribució d'electricitat estaven situats a una altura de 6 o 7 metres perquè ningú s'hi connectés de forma il·legal. La demandant defenia que això li impedia tenir control del comptador del seu consum

[36] TJUE (Gran Sala). Sentència de 16 de juliol de 2015, *CHEZ Razpredelenie Bulgaria AD* v *Komisia za zashtita ot diskriminatsia*, C-83/14, EU:C:2015:480.

elèctric i que es tractava d'una discriminació racial —encara que ella no fos d'ètnia romaní—. També en l'assumpte *Schönheit* contra *Stadt Frankfurt am Main*, de 23 d'octubre de 2003[37], el TJUE aborda el càlcul en base a una tarifa diferent entre les persones treballadores a temps parcial i les persones treballadores a temps complert[38]. Malgrat aplicar-se aquesta norma a totes les persones treballadores, afectava les dones d'una manera molt més exacerbada ja que suposaven el 88 % de les persones treballadores a temps parcial[39]. En resum, tot i estar integrat de manera predominant per dones, el grup desavantatjat també integrava homes[40].

Per mitigar aquesta problemàtica —com també la relativa al grup comparador que abordarem en el següent apartat—, s'ha plantejat emprar la prova estadística, però la jurisprudència del TJUE i el TEDH no sembla adaptada a les necessitats probatòries. Tant per part de TJUE[41], com per part de TEDH[42], s'ha recorregut anteriorment aquesta prova, però no s'ha fixat per part del TJUE ni del TEDH un criteri únic sobre quin percentatge de persones negativament afectades per la disposició, criteri o pràctica aparentment neutra ha de pertànyer al grup desavantatjat, ja que tendeixen a analitzar la situació de manera específica[43]. Tampoc s'ha establert el tant per cent de persones amb el mateix atribut que han de resultar negativament afectades en comparació a les altres que no comparteixen aquest atribut malgrat es troben en el grup desavantatjat (exemple de persones treballadores a temps parcial) per considerar que es dona un cas de discriminació indirecta (Soriano, 2021). És més, no s'han establert requisits consistents per avaluar la fiabilitat, rellevància i importància de la prova estadística.

[37] TJUE (Sala cinquena). Sentència de 23 d'octubre de 2003, *Hilde Schönheit* contra *Stadt Frankfurt am Main,* C-4/02, EU:C:2003:583.

[38] En el mateix sentit es pronuncia el TEDH en l'assumpte Isabel Elbal Moreno contra Instituto Nacional de la Seguridad Social de 22 de novembre de 2012, on estipula que constitueix discriminació indirecta per raó de sexe una legislació estatal que exigeix per aconseguir la pensió de jubilació un període de cotització major a les persones treballadores a temps parcial que a aquelles a temps complert.

[39] Vegeu també, per exemple, la sentència del TJUE (Sala Sisena) de 11 de setembre de 2003, *Erika Steinicke* contra *Bundesanstalt für Arbeit* (C-77/02) o la sentència del TJUE de 27 de maig de 2004, *Edeltraud Elsner-Lakeberg* contra *Land Nordrhein-Westfalen* (285/02).

[40] La mateixa qüestió però sentenciada a nivell nacional la trobem en la sentència 91/2019, de 3 de juliol de 2019, on el TC va declarar inconstitucional i nul un precepte que regulava el càlcul de la pensió de la jubilació de les persones treballadores a temps parcial per considerar que constituïa una discriminació indirecta per raó de sexe.

[41] TJUE. Sentència de 9 de febrer de 1999. *Regina* contra *Secretary of State for Employment*, ex parte *Nicole Seymour-Smith y Laura Pérez*. C-167/97. EU:C:1999:60. para. 61 i 62.

[42] TEDH, D.H. i altres contra la República Txeca, núm. demanda 57325/00, 2007, para. 175.

[43] Per exemple, en el cas *D.H.* i altres contra la República Txeca de 13 de novembre de 2007 (TEDH Gran Sala, Núm. demanda 57325/00), el TEDH va apreciar per primera vegada en la seva jurisprudència l'existència de discriminació indirecta al entendre que les proves psicològiques que s'empraven per determinar la intel·ligència i la idoneïtat —a la que es sotmetia a tots els alumnes per igual—, per avaluar si havien de ser apartats de la educació general i adscrits a escoles dissenyades per nens amb necessitats especials causaven discriminació indirecta per raó d'ètnia, ja que entre el 80 % i el 90 % dels infants d'ètnia romaní acabaven fora del sistema general com a conseqüència d'aquestes proves. El tribunal va prendre aquesta decisió perquè va considerar desproporcionada la presencia de menors d'ètnia romaní als centres escolars per infants amb necessitats especials ja que representaven un 50 % dels infants dels centres quan només un 2 % dels menors txecs eren d'ètnia romaní.

Sovint es delega l'establiment d'aquests requisits als tribunals nacionals, generant uns estàndards de prova divergents entre Estats de la UE (Tobler, 2008).

4.3.3. Impediments per trobar un grup comparador

En el cas del grup comparador, el Tribunal necessita que el terme de comparació sigui apropiat per determinar si el criteri afecta de manera més significativa i perjudicial a un grup —que es pugui determinar en relació a un o diversos trets comuns— que a un altre (Rey, 2019). Per exemple, en el cas *Osterreichischer Gewerkschaftsbund*, el TJUE estableix com a no comparable que les dones agafin permisos de maternitat i els homes permisos per a realitzar el servei militar[44].

Per identificar el grup comparador, destaquen dues qüestions pràctiques. Primer, amb qui pot comparar-se la persona demandant legítimament i quina argumentació o evidència es requereix per establir que els dos grups s'haurien de considerar iguals en una mateixa situació?. Com ja hem vist, les comparacions binàries son rares en els sistemes algorítmics, on es tenen en compte molts atributs. Segon, s'ha d'identificar un grup comparador concret per establir discriminació *prima facie* o és suficient un comparador hipotètic? (Wachter, Mittelstadt i Russell, 2021). Com es pot veure en els casos *Maniero*[45] o *Jyske Finans*[46], les persones demandants al·leguen haver patit discriminació indirecta per raó d'ètnia i el TJUE els demana un grup comparador concret i específic en comptes d'un general i abstracte. Tot i així, per posar un exemple —malgrat no sigui del TJUE—, el Tribunal Social Suec va sentenciar un cas relatiu a l'exigència de la llengua sueca per treballar, admetent discriminació indirecta *prima facie* contra qui no tingués la nacionalitat sueca, esdevenint el grup de persones nacionals sueques com a comparador, és a dir, un grup general (Farkas i O'Dempsey, 2011).

La legitimitat del grup comparador depèn de les circumstàncies del cas, dels arguments presentats i de les resolucions dels tribunals nacionals. Les directives antidiscriminació rarament ofereixen requisits clars que no necessitin futures interpretacions jurisprudencials, fent que la determinació de la legitimitat del grup comparador sigui contextual. De manera què, inclús quan les evidències *prima facie* es

[44] TJUE (Gran Sala). Sentència de 8 de juny de 2004, *Österreichischer Gewerkschaftsbund, Gewerkschaft der Privatangestellten* contra *Wirtschaftskammer Österreich*, C-220/02, EU:C:2004:334. Para. 64.

[45] En aquest assumpte, sentenciat el 15 de novembre de 2018, la persona demandant al·legava discriminació indirecta segons la Directiva d'igualtat racial que limitava la possibilitat d'obtenir una beca alemanya a aquelles persones que havien aprovat un examen final exigible a les persones estudiants de dret de la universitat alemanya. La persona demandant, havent cursat el grau en dret en una universitat d'un altre país al·legava que aquest criteri perjudicava a persones en base a l'ètnia. El TJUE va concloure que era necessària una comparació concreta i no una general i abstracta (para. 48).

[46] En aquest assumpte, sentenciat el 8 de desembre de 2005, la persona demandant al·legava que la llei danesa que requeria identificació addicional per ciutadans de països tercers per obtenir un préstec per comprar un cotxe causava discriminació indirecta en base a l'origen. El TJUE va establir que era necessària una comparació especifica i concreta i no una comparació abstracta i general (para. 32). El TJUE també argumenta que el país de naixença no justifica, per sé, la pertinença a un grup ètnic (para. 20) i que no es pot presumir que cada estat té un únic grup ètnic (para. 21). En aquest cas la persona demandant hauria d'haver demostrat que la nacionalitat és una representació fiable de l'ètnia.

poden aconseguir, pot ser un repte per als tribunals avaluar el test de necessitat, proporcionalitat i justificació objectiva que cal fer en la discriminació indirecta[47].

Com hem vist en aquesta secció, el funcionament dels sistemes algorítmics essencialment col·lectiu (la discriminació es crea en base al tractament de dades de múltiples persones que integren diferents —i també múltiples— col·lectius, i no en base a la persona en si) genera dificultats afegides per a l'establiment tant d'un grup desavantatjat, com d'un grup comparador i, per tant, limita la protecció —que actualment parteix d'una perspectiva essencialment individual— envers la discriminació algorítmica.

5. CONCLUSIONS

En aquest article hem recollit els que considerem que són els principals límits de la normativa antidiscriminació europea davant les discriminacions algorítmiques i els hem dividit en tres eixos. Arran de la detecció d'aquests límits hem pretès senyalar algunes línies que els poden mitigar.

En la Secció 2, estudiem com mitigar el fet que la protecció europea actual sigui excessivament depenent dels atributs concrets a protegir, tenint en compte la capacitat que caracteritza els sistemes algorítmics, per analitzar, relacionar i inferir dades personals. Seguint a Xenidis (2020), s'ha d'apostar per treballar des d'una perspectiva interseccional, que a banda de posar en dubte que la normativa antidiscriminatòria no tingui una perspectiva més oberta dels atributs a protegir, aconsegueix copsar millor la capacitat dels sistemes algorítmics per discriminar en base a múltiples atributs.

En la Secció 3, hem analitzat en quins aspectes la discriminació algorítmica té dificultats d'encaix, tant en el marc de la discriminació directa com en el marc de la discriminació indirecta. Donat que les subtileses de les noves formes de discriminació algorítmica inclinen a pensar que s'enquadra millor en la forma indirecta de discriminació, es fa palesa llavors la necessitat de repensar el sistema de justificació general objectiva, que dona molta llibertat a qui presumptament discrimina. També, com la capacitat dels sistemes algorítmics d'inferir dades personals a través d'indicadors, qüestiona la dualitat —en els casos en que hi ha una superposició del significat de l'indicador amb l'atribut protegit— dels conceptes de discriminació directa i discriminació indirecta.

Finalment, en la Secció 4, interpretant que la persona és conscient de la discriminació (havent salvat l'opacitat inherent a aquesta tipologia de presa de decisió) i pot identificar qui es responsable de l'acció discriminatòria, la perspectiva individual de la protecció europea, la manca d'accés a informació i la personalització que

[47] En aquest sentit, vegeu també: TJUE. Sentència de 10 de juliol de 2008, *Centrum voor gelijkheid van kansen en voor racismebestrijding* contra *Firma Feryn NV*, C-54/07, EU:C:2008:397; TJUE, 25 d'abril de 2013, *Asociația Accept* contra *Consiliul Național pentru Combaterea Discriminării*, C-81/12, EU:C:2013:275; TJUE, Sentència de 23 d'abril de 2020, *NH* contra *Associazione Avvocatura per i diritti LGBTI - Rete Lenford*, C 507/18, EU:C:2020:289.

proporcionen els sistemes algorítmics, ocasionen la quasi impossibilitat de trobar un grup de persones desavantatjades i un grup de persones amb qui comparar-se, per tant esdevé molt complicat assumir l'exigència dels tribunals per l'establiment d'un cas *prima facie*. Hem plantejat, donada la situació actual, la possibilitat d'automatitzar la inversió de la càrrega de la prova en els casos de discriminació algorítmica per manca d'informació —entenent que existeix una asimetria de poder en aquests casos que suposa *de facto*, la més que probable negativa de qui empra el sistema algorítmic a donar informació necessària per invertir la càrrega de la prova—. També considerem que caldria treballar per generar eines que duguin cap a una protecció col·lectiva o que, si més no, donin accés a dades grupals que ajudin a entendre i possibilitin discutir casos individuals, ja que només amb dades individuals i una perspectiva individual, la protecció que actualment brinda la normativa antidiscriminatòria europea no esdevé un sistema efectiu per abordar les discriminacions algorítmiques.

BIBLIOGRAFIA

Allhutter D., Cech F., Fischer F., Grill G. & Mager A. (2020). Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective. *Frontiers Big Data*, *3*(5). https://www.frontiersin.org/articles/10.3389/fdata.2020.00005/full

Barocas, S. i Selbst, A. D. (2016). Big Data's Disparate Impact, *Cal. Law Rev*., 104, 671-732. https://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf

Crawford, K. (2021). *Atlas of AI*. Yale University Press.

Eguíluz, J.A. (2020). Desafíos y retos que plantean las decisiones automatizadas y los perfilados para los derechos fundamentales. *Estudios de Deusto*, *68*(2), 325-367. https://doi.org/10.18543/ed-68(2)-2020pp325-367

Farkas, L. i O'Dempsey, D. (2011). *How to present a discrimination claim: handbook on seeking remedies under the EU non-discrimination directives*. European Commission. https://op.europa.eu/s/wAbr

Farkas, L i O'Farrel, O. (2015). *Reversing the burden of proof: practical dilemmas at the European and national level. European Network of legal experts in the Non-discrimination field*. European Commission. https://op.europa.eu/s/wAbq

Gerards, J. i Xenidis, R. (2021). *Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non-Discrimination Law*. European Commission. https://op.europa.eu/s/wAbs

Gillis, T.B. i Spiess, J.L. (2019) Big Data and Discrimination. *University of Chicago Law Review*, *86*(2), 459-487. https://chicagounbound.uchicago.edu/uclrev/vol86/iss2/4

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. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3164973

Kullmann, M. (2018). Platform Work, Algorithmic Decision Making, and EU Gender Equality Law. *International Journal of Comparative Labour Law and Industrial Relations, 34*(1), 1-21. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3195728

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. https://doi.org/10.1177/0967010614544204

Páez, A. (2021). Negligent Algorithmic Discrimination. *Law and Contemporary Problems*, *84* (3). 19-33. https://scholarship.law.duke.edu/lcp/vol84/iss3/3

Requena, M. (2011). La tutela judicial del principio general de igualdad de trato en la Unión Europea: una jurisprudencia expansiva basada en una jerarquía de motivos discriminatorios. *Revista de Derecho Comunitario Europeo*, 40, 767-793. https://recyt.fecyt.es/index.php/RDCE/article/view/39481

Rey, F. (2019) *Derecho Antidiscriminatorio*. Thompson- Aranzadi.

Ringelheim, J. (2019). The Burden of Proof in Antidiscrimination Proceedings. A Focus on Belgium, France and Ireland. *European Equality Law Review*, (2). https://www.equalitylaw.eu/downloads/5005-european-equality-law-review-2-2019-pdf-3-201-kb

Rivas-Vallejo, P. (2021). *Discriminación algorítmica: detección, prevención y tutela*. XXXI Jornades Catalanes de Dret Social ("Treball, discriminació i COVID"). http://www.iuslabor.org/jornades-i-seminaris/ponencies/any-2021/

Roig, A. (2020). *Las garantías frente a las decisiones automatizadas. Del Reglamento General de Protección de Datos a la gobernanza algorítmica*. J.M. Bosch Editor.

Soriano, A. (2021). Decisiones automatizadas y discriminación: aproximación y propuestas generales. *Revista General de Derecho Administrativo*, 56. http://laadministracionaldia.inap.es/noticia.asp?id=1511706

Tobler, C. (2008). *Limits and potential of the concept of indirect discrimination. Technical report*. European Commission. https://op.europa.eu/s/wAn1

Wachter, S., Mittelstadt, B. i Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law and Security Revue, 41.* https://doi.org/10.1016/j.clsr.2021.105567

Weatherall, K. G. i Caetano, T. (2020). *Submission in response to Australian Human Rights Commission Discussion Paper, 'Human Rights and Technology', April 2020*. Sydney Law School Research Paper, 20/27. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3609578

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

Xenidis, R. (2017). *Shaking the normative foundations of EU equality law: evolution and hierarchy between market integration and human rights rationales.* EUI Working Papers Law, 2017/04. https://cadmus.eui.eu/bitstream/handle/1814/45489/LAW_2017_04.pdf?sequence=3

Xenidis, R. i Senden, L. (2020). EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination. In: Ulf Bernitz i altres (eds). *General Principles of EU law and the EU Digital Order,* 151-182. https://cadmus.eui.eu/handle/1814/65845

Zuiderveen, F. (2018). *Discrimination, artificial intelligence, and algorithmic decision-making*. Consell d'Europa. https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73