

Anna Capellà Ricart

**Discriminació
algorítmica:
garanties des de la
normativa europea
antidiscriminació,
l'RGPD i el Reglament
de la intel·ligència
artificial**

Aquest llibre fa una aproximació a la principal normativa europea que regula els resultats discriminatoris que poden causar les decisions preses mitjançant sistemes algorítmics. Es fa una anàlisi sobre l'adequació de la normativa antidiscriminació europea per abordar les característiques particulars de la discriminació algorítmica, estudiant també les garanties per la protecció del dret a la no discriminació que aporta el Reglament General de Protecció de Dades i el Reglament de la intel·ligència artificial. Aquestes dues últimes normatives tenen la capacitat de suplir certes mancances que presenta la normativa antidiscriminació europea en el context específic de la discriminació algorítmica, atès que aquesta presenta elements que no trobem en la discriminació tradicional (p. ex. anàlisi massiu de dades, algoritmes, multiplicitat d'actors que intervenen en la presa de decisió). L'estudi se centra en l'accés a la informació sobre les decisions automatitzades que pugui ser útil per provar casos de discriminació algorítmica, en la consecució de sistemes algorítmics menys esbiaixats i per tant amb menys probabilitats de generar resultats discriminatoris i en l'establiment de drets per a les persones que presumptivament s'hagin vist discriminades a fi de promoure una protecció més eficient.

Anna Capellà Ricart

**Discriminació algorítmica:
garanties des de la normativa
europea antidiscriminació,
l'RGPD i el Reglament de
la intel·ligència artificial**

Barcelona
2024



BOSCH EDITOR

© OCTUBRE 2024 ANNA CAPELLÀ RICART

© OCTUBRE 2024



Librería Bosch, S.L.

<http://www.jmboscheditor.com>

<http://www.libreriabosch.com>

E-mail: editorial@jmboscheditor.com

Qualsevol forma de reproducció, distribució, comunicació pública o transformació d'aquesta obra només es pot fer amb l'autorització dels seus titulars, llevat d'excepció prevista per la llei. Adreceu-vos a CEDRO (*Centro Español de Derechos Reprográficos*) si necessiteu fotocopiar o escanejar algun fragment d'aquesta obra (www.conlicencia.com; 91 702 19 70 / 93 272 04 45).

ISBN paper: 978-84-10448-16-2

ISBN digital: 978-84-10448-17-9

D.L.: B 17579-2024

Disseny portada i maquetació: CRISTINA PAYÁ  +34 672 661 611

Printed in Spain – Impres a Espanya



Índex

Introducció.....	13
------------------	----

Capítol I

Possibilitats de la regulació europea antidiscriminació per abordar la discriminació algorítmica.....	21
---	----

1. Llistes obertes d'atributs protegits i homogeneïtzació de la protecció.....	25
1.1. Llistes obertes d'atributs protegits.....	26
1.2. Homogeneïtzació de la protecció de les directives antridiscriminació	32
2. Reconeixement del caràcter grupal en la discriminació algorítmica.....	35
2.1. Reconeixement de la discriminació estructural	36
2.2. Reconeixement de la discriminació interseccional.....	39
3. Discriminació directa i discriminació indirecta	45
3.1. Elements d'interès de cada tipologia de discriminació.....	45
3.2. Discriminació per associació i per error	57
4. Possibilitats de millora de la prova	70

4.1. No necessitat de condició de víctima (cas Feryn) i suficiència d'un dany hipotètic (cas Accept i Rete Lenford)	71
4.2. Accés a la informació: Cas Meister	75

Capítol II

Aportacions de l'RGPD a la protecció contra la discrimina- ció algorítmica	81
1. Possibilitats de l'RGPD com a sistema de governança col·laborativa per abordar la discriminació algorítmica	85
1.1. L'RGPD com a sistema de governança col·labo- rativa	85
1.2. Condicions per l'efectivitat d'un règim de go- vernança col·laborativa	88
2. Qualitat de les dades per abordar els biaixos algorít- mics	95
2.1. L'exactitud	95
2.2. Drets de les persones sobre les seves dades: accés, supressió, rectificació, limitació del ter- mini de conservació i portabilitat	98
3. Informació útil i senzilla per fer front a la discrimi- nació algorítmica	104
3.1. El dret a una explicació en la protecció envers la discriminació	105
3.2. Contingut de la informació individual neces- sària per iniciar un plet de discriminació	112

3.3.	Política de transparència de l'RGPD	118
4.	Responsabilitat proactiva plantejada en l'RGPD.....	120
4.1.	Sistemes d'autoregulació: Codis de conducta, certificacions, segells i marques	121
4.2.	Igualtat des del disseny i per defecte.....	126
4.3.	De les avaluacions d'impacte de protecció de dades a les avaluacions d'impacte algorítmiques	131
5.	Vessant col·lectiva de la protecció envers la discriminació: Article 80 RGPD	144
5.1.	Article 80.2 RGPD: Representació sense mandat exprés	145
5.2.	Possibilitats de l'article 80.2 RGPD per redreçar la discriminació algorítmica	148
5.3.	Dificultats per l'exercici de l'article 80 en relació amb la discriminació algorítmica	150

Capítulo III

	Discriminació algorítmica i Reglament de la Intel·ligència Artificial.....	155
1.	El principi de no-discriminació en el RIA.....	158
1.1.	Un Reglament per assegurar el bon funcionament del mercat intern de la UE	160
1.2.	Drets de les persones afectades pels sistemes d'IA.....	163

2.	Protecció de la discriminació algorítmica segons el grup de risc en el qual estigui classificat el sistema d'IA.....	165
2.1.	Sistemes prohibits.....	168
2.2.	Sistemes de risc alt	170
2.3.	Sistemes que generen riscos que comporten requisits específics de transparència	173
2.4.	Sistemes d'IA d'ús general amb risc sistèmic..	174
3.	Elements del RIA rellevants per la protecció envers la discriminació	175
3.1.	Disseny i qualitat de les dades	175
3.2.	Documentació tècnica relativa als sistemes de risc alt.....	182
3.3.	Supervisió humana dels sistemes de risc alt...	187
3.4.	Avaluació d'impacte relativa als drets fonamentals per als sistemes d'IA d'alt risc.....	189
3.5.	Base de dades pública.....	192
3.6.	Espais controlats de proves per a la IA.....	194
4.	No discriminació i estandardització de sistemes d'IA.....	196
4.1.	Trasllat del principi de no-discriminació als estàndards creats per les organitzacions europees d'estandardització.....	197
4.2.	Avaluació de conformitat.....	207

4.3. Marcatge “CE” de conformitat.....	210
4.4. Seguiment posterior a la comercialització o posada en servei del sistema d’IA	211
Conclusions	215
Bibliografia	221
Bibliografia web	241
Annex.....	243



Introducció

La discriminació es podria definir com el tracte desfavorable, d'inferioritat o excloent que es dona a una persona o un col·lectiu per raó d'una condició o circumstància de tipus personal o social (p. ex. edat, sexe o orientació sexual)¹. Aquest tracte desfavorable, d'inferioritat o excloent es pot donar a través de decisions *únicament* automatitzades², que són aquelles que es basen només en mitjans tecnològics sense la participació de l'ésser humà³. En el cas que a part dels mitjans tecnològics, en les decisions també hi participin persones, s'anomenaran decisions automatitzades. Aquest tipus de decisions poden ser preses mitjançant intel·ligència artificial (també l'anomenarem al llarg del treball, IA), que s'entén com la capacitat d'un ordinador o un robot controlat per ordinador per

- 1 En aquest treball anomenarem a aquestes circumstàncies de tipus personal o social protegides per la normativa com a atributs protegits, categories sensibles, característiques protegides, motius protegits o categories protegides.
- 2 Al llarg del treball quan ens referim a decisions automatitzades, únicament automatitzades, sistemes algorítmics o sistemes d'IA en referim als casos en que aquestes s'apliquen en contextos socials i poden produir discriminació.
- 3 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, 17/EN, WP 251, Brussels, 6 February 2018, pàg. 8. Arran de l'entrada en vigor del Reglament General de Protecció de Dades, aquest grup de treball europeu independent s'anomena Comitè Europeu de Protecció de Dades (CEPD), en anglès, European Data Protection Board (EDPB). Aquest canvi es deu a que el Grup de Treball de l'Article 29 interpretava la Directiva 95/46/CE, que va quedar derogada amb l'entrada en vigor de l'RGPD.

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 passades⁴.

Quan la discriminació es produeix mitjançant decisions automatitzades, que operen a través d'algoritmes o sistemes algorítmics, s'anomenarà discriminació algorítmica, que de la mateixa manera que la discriminació tradicional⁵, és el resultat de prejudicis i estereotips sobre col·lectius determinats de persones.

Per tant, les decisions automatitzades, tal com les decisions no automatitzades, causen situacions discriminatòries. L'automatització brinda eficàcia, celeritat, reducció de costos i capacitat per processar i analitzar una quantitat de dades que la ment humana no pot abastar. Tot i així, una màquina, inclús en el cas de les decisions automatitzades, segueix unes directrius donades, per tant, l'elecció sobre quines dades s'utilitzen, com es tracten i les conclusions a les que s'ha d'arribar o els objectius que té l'acció, estan, en primera instància, estipulats per persones. La discriminació que pot produir un sistema d'intel·ligència artificial pot venir ocasionada –a més d'en base a apreciacions humanes– a causa d'errors i biaixos en les bases de dades d'entrenament dels algoritmes (tant en la construcció com després de la construcció de la base de dades)⁶,

4 Artificial Intelligence. Britannica, 2023. Disponible a: <https://www.britannica.com/technology/artificial-intelligence>. Darrera visita el 17 de juliol de 2024.

5 Anomenarem tradicional aquella discriminació que no es produeix mitjançant decisions automatitzades.

6 SURESH, H. i GUTTAG, J., "Framework for Understanding Unintended Sources of Harm Throughout the Machine Learning Life Cycle", EAAMO 2021: ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, 2021, pàg. 4-6.

de la selecció i pes atorgat a les variables⁷, de la selecció de les característiques de classificació de les dades⁸, de l'obtenció de dades personals mitjançant correlacions amb altres dades⁹, de l'assumpció o l'associació de les dades amb grups socials discriminats¹⁰ o de la perpetuació d'estereotips en els quals es basen les estructures socials de discriminació¹¹. L'existència de biaixos algorítmics, afegits al grau de personalització al qual poden arribar els sistemes algorítmics atesa l'anàlisi massiva de dades, la magnificació dels efectes discriminatoris i l'opacitat inherent a molts sistemes d'IA, fan que la discriminació ocasionada mitjançant sistemes d'intel·ligència artificial pugui tenir molta incidència negativa en la vida de les persones, ocasionant riscos pels seus drets i les seves llibertats fonamentals.

L'estudi que aquí presentem recull els debats principals que hi ha en relació amb la capacitat de la normativa antidiscriminació europea per abordar la discriminació algorítmica i les solucions plantejades (Capítol I), i les garanties que es poden extreure tant

-
- 7 SORIANO ARNANZ, A., "Decisiones automatizadas y discriminación: aproximación y propuestas generales", *Revista General de Derecho Administrativo*, núm. 56, 2021, pàg. 11.
 - 8 BAROCAS, S. i SELBST, A. D., "Big Data's Disparate Impact", *California Law Review*, vol. 104, 2016, pàg. 688.
 - 9 HUERGO LORA, A., "Una aproximación a los algoritmos desde el derecho administrativo", a HUERGO LORA, A.J., (dir.) i DÍAZ GONZÁLEZ, G. M. (coord.), *La regulación de los algoritmos*, Pamplona, Aranzadi, 2020, pàg. 46 i 47.
 - 10 GERARDS, J. i XENIDIS, R., *Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non-Discrimination Law*, European Commission, Brussels, pàg. 68.
 - 11 GERARDS, J. i ZUIDERVEEN, F., "Protected Grounds and the System of non-discrimination law in the context of algorithmic decision-making and artificial intelligence", *Colorado Technology Law Journal*, vol. 20, núm. 1, 2022, pàg. 15.

del Reglament General de Protecció de Dades europeu¹² (d'ara en endavant, RGPD) (Capítol II) com del Reglament de la intel·ligència artificial europeu¹³ (d'ara en endavant, RIA), recentment aprovat pel Parlament Europeu (Capítol III), que al tenir en compte mesures tècniques poden suplir certes mancances que presenta la normativa antidiscriminació europea en el context específic de les decisions automatitzades. L'estudi se centra en l'accés a la informació sobre les decisions automatitzades que pugui ser útil per provar casos de discriminació algorítmica, en la consecució de sistemes algorítmics menys esbiaixats i per tant amb menys probabilitats de generar resultats discriminatoris i en l'establiment de drets de les persones presumptivament discriminades mitjançant aquests sistemes. En l'Annex s'adjunta un esquema de la interacció entre les garanties de cada normativa que poden ser útils per abordar la discriminació algorítmica.

En el pla semàntic, com ens centrarem en el debat jurídic i no en la precisió tècnica, utilitzarem, per referir-nos al mateix concepte, les expressions “sistemes d'intel·ligència artificial”, “sistemes algorítmics”, “sistemes de presa de decisions mitjançant algoritmes” i “decisiones automatitzades mitjançant sistemes algorítmics”. També utilitzarem frases com “efectes de la intel·ligència artificial” tal com si la intel·ligència artificial fos una entitat que actués per si mateixa.

12 Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos). DOUE, L 119/1, 4.5.2016.

13 Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). DOUE, sèrie L, 12.7.2024, 1-144.

Així i tot, els sistemes d'intel·ligència artificial no es creen espontàniament i no tenen per ells mateixos efectes adversos en els drets humans, sinó que els tenen en tant s'implementen i s'apliquen a la interacció humana.

També cal esmentar que les decisions automatitzades que empren intel·ligència artificial es poden dur a terme mitjançant algoritmes no predictius o mitjançant algoritmes predictius. Seguirem l'exemple de Huergo Lora a "*La regulación de los algoritmos*", basat en el sistema d'acreditació de professorat universitari. Qui realitza la norma decideix, fent un judici de valor, quines són les característiques d'un/a professor/a ideal i puntua les persones aspirants en el grau en què les posseeixen. Un algoritme no predictiu calcularà les puntuacions en base a les dades curriculars (publicacions, docència, etc.). En canvi, un algoritme predictiu, analitzarà les dades del professorat considerat millor o més productiu d'acord amb els criteris prèviament establerts; seguidament, determinarà quines característiques formatives o curriculars estan associades a aquest nivell d'excel·lència; i finalment, identificarà d'entre les persones aspirants, aquelles que reuneixen aquestes característiques i, per tant, es pot preveure que arribaran al nivell establert¹⁴.

Els algoritmes predictius són la base de l'aprenentatge automàtic. Però, que és l'aprenentatge automàtic (en tant que eina que pot ser utilitzada en la presa de decisions automatitzades)? Alpaydin observà que els sistemes que programaven els investigadors per resoldre qüestions tenien dos desavantatges, en primer lloc, que les normes lògiques en els sistemes no sempre encaixaven la realitat canviant del món (en la vida quotidiana les coses no són certes o falses pròpiament); i en segon lloc, les persones investigadores havien de proveir el coneixement (les respostes) per introduir-lo en

14 HUERGO LORA, A.J., "Una aproximación a los algoritmos...", op. cit., 2020, pàg. 38.

el sistema, un procés que costava molts diners i temps¹⁵. Arran de l'aprenentatge automàtic, el coneixement va passar de proveir-se per part de persones expertes a ser extret pel propi sistema en base a exemples i experiència en forma de dades (concretant-li una tasca i alimentant-lo amb una gran quantitat de dades per usar-les com a exemples de com aquesta tasca es pot aconseguir o de quines dades extreure'n patrons)¹⁶. Els recents avenços de la intel·ligència artificial es deuen, sobretot, al desenvolupament d'aquesta tècnica. Una subclasse de l'aprenentatge automàtic és l'aprenentatge profund (Deep learning). Dins d'aquest es pot distingir entre l'aprenentatge supervisat, és a dir, quan les dades d'entrada que utilitza el sistema són etiquetades per persones, i l'aprenentatge no supervisat, és a dir, quan les dades d'entrada no són etiquetades per persones i és un algorisme que realitza la seva pròpia classificació¹⁷.

Entre les dades que s'utilitzen per entrenar els sistemes n'hi ha que en aparença són neutres però revelen atributs personals. Aquestes dades neutres es poden anomenar indicadors (en anglès, proxy), i s'entenen per part de les persones que programen els sistemes algorítmics com a elements que estan a prop i per tant poden servir com a aproximació a altres elements als quals no tenen accés directe (aquests elements als quals no tenen accés directe poden ser els atributs personals especialment protegits per la normativa). D'aquesta manera, malgrat el sistema no empri explícitament atributs personals pot acabar generant –arran de l'ús d'indicadors– resultats discriminatoris.

15 ALPAYDIN, E., *Machine learning: the new AI*, Massachusetts, The MIT Press, 2016, pàg. 51.

16 THE ROYAL SOCIETY, *Machine learning: the power and promise of computers that learn by exemple*, London, April 2017, pàg.19.

17 OBSERVATORI D'ÈTICA EN INTEL·LIGÈNCIA ARTIFICIAL DE CATALUNYA (OEIAC), *Intel·ligència artificial, ètica i societat: Una mirada i discussió a través de la literatura especialitzada i d'opinions expertes*, Girona, 2021, pàg. 16 i 17.

Finalment, és rellevant diferenciar entre els conceptes de discriminació directa i discriminació indirecta. La discriminació directa és aquella situació en la que 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 gènere)¹⁸. Es pot justificar la discriminació directa però només en base a certs criteris taxats¹⁹ i la jurisprudència del Tribunal de Justícia de la Unió Europea (d'ara en endavant, TJUE) ha admès que no sigui necessari una persona en situació comparable per poder apreciar-la. En canvi, la discriminació indirecta es podria definir com aquella situació en la qual una pràctica que aparentment és neutra acaba situant en una posició de desavantatge particular a persones que es poden identificar amb alguna característica protegida en comparació amb persones que no ostenten la categoria protegida²⁰. Es pot justificar la discriminació indirecta objectivament (en conseqüència les causes de justificació són obertes) i per apreciar-la requereix un grup de persones que s'hagi vist desavantatjat per la pràctica respecte a un altre grup, que actuarà com a comparador.

18 REY MARTÍNEZ, F., *Derecho Antidiscriminatorio*, Navarra, Aranzadi, 2023, pàg. 70.

19 Article 3.4 de la Directiva 2000/78/CE; Article 4.1 de la Directiva 2000/78/CE. També, de la mateixa manera, trobem aquesta exclusió en l'article 14.2 de la Directiva 2006/54/CE i l'article 4 de la Directiva 2000/43/CE; Article 4.2 de la Directiva 2000/78/CE; Article 6 de la Directiva 2000/78/CE.

20 TJUE (Sala Tercera), sentència de 24 de febrer de 2022, C-389/20, CJ contra Tesorería General de la Seguridad Social (TGSS), par. 40.

Introducció

Capítol I

Possibilitats de la regulació europea antidiscriminació per abordar la discriminació algorítmica

Capítol II

Aportacions de l'RGPD a la protecció contra la discriminació algorítmica

Capítulo III

Discriminació algorítmica i Reglament de la Intel·ligència Artificial

Conclusions

Bibliografia

Bibliografia web

Annex

Anna Capellà i Ricart

és Doctora en Dret per la Universitat Autònoma de Barcelona (2023). Actualment treballa a l'Institut de Dret i Tecnologia (IDT) de la Universitat Autònoma de Barcelona. Les seves línies d'investigació són la perspectiva antissubordinatòria del dret antidiscriminatori, la garantia constitucional del dret a la igualtat i a la no discriminació i la garantia col·lectiva de drets en el context digital.