## INTELIGENCIA ARTIFICIAL Y SECTOR PÚBLICO



### Retos, límites y medios

EDUARDO GAMERO CASADO Director

FRANCISCO L. PÉREZ GUERRERO Coordinador



### INTELIGENCIA ARTIFICIAL Y SECTOR PÚBLICO

Retos, límites y medios

#### **EDUARDO GAMERO CASADO**

Director

#### FRANCISCO L. PÉREZ GUERRERO

Coordinador









Programa Operativo FEDER 2014-2020

Proyecto de investigación UPO-1381574, "Inteligencia artificial y Derecho Administrativo"

Red DAIA Derecho Administrativo e Inteligencia Artificial http://reddaia.org/

#### tirant lo blanch

Valencia, 2023

#### Copyright ® 2023

Todos los derechos reservados. Ni la totalidad ni parte de este libro puede reproducirse o transmitirse por ningún procedimiento electrónico o mecánico, incluyendo fotocopia, grabación magnética, o cualquier almacenamiento de información y sistema de recuperación sin permiso escrito de los autores y del editor.

En caso de erratas y actualizaciones, la Editorial Tirant lo Blanch publicará la pertinente corrección en la página web www.tirant.com.

© Eduardo Gamero Casado (Dir.) Francisco L. Pérez Guerrero (Coord.)

© TIRANT LO BLANCH

EDITA: TIRANT LO BLANCH

C/ Artes Gráficas, 14 - 46010 - Valencia

TELFS.: 96/361 00 48 - 50

FAX: 96/369 41 51 Email:tlb@tirant.com

www.tirant.com

Librería virtual: www.tirant.es DEPÓSITO LEGAL: V-1554-2023

ISBN: 978-84-1169-063-8

MAQUETA: Disset Ediciones

Si tiene alguna queja o sugerencia, envíenos un mail a: atencioncliente@tirant.com. En caso de no ser atendida su sugerencia, por favor, lea en www.tirant.net/index.php/empresa/politicas-de-empresa nuestro procedimiento de quejas.

Responsabilidad Social Corporativa: http://www.tirant.net/Docs/RSCTirant.pdf

6 CONCLUSIONES	249
BIBLIOGRAFÍA CITADA	251
Sección Segunda: Los derechos y la legislación administrativa con	no
límites a la IA	
0 / 1 ****	
Capítulo VII	
Discriminación, sesgos e igualdad de la inteligencia artificial en el sector pú LORENZO COTINO HUESO	blico
LORENZO COTINO HUESO	
1 SESGOS Y DISCRIMINACIÓN ALGORÍTMICA ¿DE QUÉ HABLA-	
MOS? Y ALGUNOS EJEMPLOS	260
1.1 Sesgos, errores y discriminación algorítmica y la conveniencia	
de abordarlos conjuntamente	260
1.2 Algunos ejemplos de discriminación algorítmica pública	262
2 ¿POR QUÉ SE GENERAN LA DISCRIMINACIÓN Y LOS SESGOS	
ALGORÍTMICOS?	269
3 ¿RESPECTO DE QUÉ IA PÚBLICA ES EXIGIBLE LA NO DISCRI-	_0,
MINACIÓN?	272
3.1 La igualdad es exigible respecto de los sistemas de IA públicos,	2/2
sus usos y sus efectos, aunque resulte difícil su categorización	272
3.2 El grado de legalidad exigible para diseño, implantación y uso	2/2
del sistema de IA público	276
4 ¿CUÁNDO EL USO PÚBLICO DE UN SISTEMA DE IA TIENE IM-	270
PACTO Y SUFICIENTE RELEVANCIA JURÍDICA Y PUEDE CONS-	
TITUIR UNA DISCRIMINACIÓN ALGORÍTMICA?	277
4.1 Los criterios y variables para determinar el impacto, nivel de	2//
riesgo y relevancia jurídica de los sistemas algorítmicos públicos	277
4.2 ¿Cuándo el sesgo y el error algorítmico es relevante para consti-	2//
tuir una discriminación? Datos estadísticos en discriminaciones	
indirectas y la necesidad de "recalibrarlos" a la baja para la IA	
pública	283
5 DISCRIMINACIONES ALGORÍTMICAS POR ERROR, DIRECTAS,	
INDIRECTAS, INTERESECCIONALES Y ACCIONES POSITIVAS	
ALGORÍTMICAS	285
5.1 La discriminación o sesgo algoritmo será en la mayoría de su-	203
puestos "discriminación por error"	286
5.2 Discriminaciones algorítmicas directas, interseccionales y por	_00
asociación	286
5.3 La complejidad de las discriminaciones algorítmicas indirectas	287
5.4 Acciones positivas algorítmicas, deber de diferenciación y mo-	
dulaciones y compensaciones para la minimización de la discri-	
minación	289

14 Índice

6 ¿CUÁNDO UN SISTEMA ALGORÍTMICO PÚBLICO QUE NO TIE- NE EN CUENTA CIRCUNSTANCIAS ESPECIALMENTE SOSPE- CHOSAS ES DISCRIMINATORIO?	2
<ul> <li>6.1 La igualdad "en" la ley o norma que habilite el uso del sistema de IA y el control laxo de la posibilidad de establecer diferencias de tratos razonables y no arbitrarios</li></ul>	29:
7 ¿CUÁNDO UN SISTEMA ALGORÍTMICO PÚBLICO QUE SÍ TIENE EN CUENTA CIRCUNSTANCIAS ESPECIALMENTE SOSPECHOSAS ES DISCRIMINATORIO?	2:
<ul> <li>7.2 En la mayoría de casos será necesaria una ley de calidad y con garantías que habilite a que un sistema de IA pueda diferenciar en razón de datos especialmente sospechosos de discriminación. Y no la hay</li></ul>	2:
nalmente estricto	30
8 ZONAS GRISES. GARANTÍAS ANTE EL USO POR EL SISTEMA DE IA PÚBLICO DE "PROXIES" Y GRUPOS ALGORÍTMICOS 8.1 Tratamientos diferentes a partir de "proxies" o datos afines o	3
indirectos de grupos especialmente protegidos	3(
9¿CÓMO PROBAR LA DISCRIMINACIÓN ALGORÍTMICA PÚBLI- CA? INVERSIÓN GENERAL DE LA CARGA DE LA PRUEBA Y LA	,
NECESIDAD DE LA TRANSPARENCIA Y EXPLICABILIDAD 9.1 Los indicios que debe aportar la persona afectada por el sistema de IA público. Inversión general de la carga de la prueba y sus	3
9.2 Lo que debe probar el proveedor o el usuario del sistema de IA	3
público	3
MINACIÓN ALGORÍTMICA PÚBLICA?	3

11 LA (CRECIENTE) OBLIGACIÓN DE DISEÑO Y ESTUDIOS DE IMPACTO FRENTE A LA DISCRIMINACIÓN ALGORÍTMICA DE
LOS SISTEMAS DE IA PÚBLICOS
11.2 Los usos públicos de IA de alto riesgo habrán de cumplir, además, los deberes del RIA
11.3 La extensión de estudios de impacto y la obligatoriedad de medidas para la no discriminación de los sistemas de IA públi-
cos
12 Y SI LOS REMEDIOS NO VALEN, ¿QUÉ PASA CUANDO SE DETECTA UNA DISCRIMINACIÓN ALGORÍTMICA PÚBLICA? 12.1 Obligaciones de mitigación, rectificación, corrección e incluso cese de uso del sistema de IA público
12.2 Nulidad, anulabilidad e inderogabilidad de los sistemas de IA público y sus efectos
12.3 Los efectos de la discriminación algorítmica en el contexto de la contratación administrativa
13 ¿QUIÉNES PUEDEN REACCIONAR ANTE UNA DISCRIMINA- CIÓN ALGORÍTMICA Y QUÉ AUTORIDADES DEBEN ACTUAR?
BIBLIOGRAFÍA CITADA
Capítulo VIII
Las singularidades del tratamiento de datos de carácter personal en entorno inteligencia artificial en el sector público  JULIÁN VALERO TORRIJOS
1 EL USO DE DATOS PERSONALES POR LAS ADMINISTRACIONES PÚBLICAS Y LA INTELIGENCIA ARTIFICIAL
2 EL CONTEXTO TECNOLÓGICO Y REGULATORIO
de estudio: el planteamiento del RGPD
la actividad administrativa
DE CARÁCTER PERSONAL

#### Capítulo VII

## Discriminación, sesgos e igualdad de la inteligencia artificial en el sector público

#### LORENZO COTINO HUESO<sup>1</sup>

Catedrático de Derecho Constitucional Universitat de Valencia

Resumen: Se definen y se presentan algunos ejemplos de sesgos y discriminación algorítmica y se exponen las causas habituales por los que se generan. Asimismo, se exponen los criterios de cuándo los usos públicos de IA pueden tener impacto y relevancia como discriminatorios. Se proyectan al ámbito de los algoritmos los diferentes conceptos de discriminación, subrayando que por lo general se tratará de discriminación "por error". Se exponen los casos en los que puede ser útil acudir a la discriminación "indirecta", se explican las "acciones positivas algorítmicas", combinadas con las acciones de diferenciación y las modulaciones y compensaciones para minimizar la discriminación. Sobre estas bases, los criterios de cuándo un sistema algorítmico público es discriminatorio son detallados. En primer lugar, cuando NO se trata de discriminación especialmente prohibida, se afirma que el control jurídico se superpone en particular con el de arbitrariedad y las exigencias de motivación. Esto sucede en especial en los usos de sistemas IA para ejercer potestades discrecionales. En segundo lugar, cuando un sistema algorítmico público SÍ tiene en cuenta circunstancias especialmente sospechosas, se sostiene que por lo general las garantías se superponen a las garantías de los tratamientos de datos especialmente protegidos. Así, para que un algoritmo diferencie por circunstancias prohibidas se impone una ley de calidad y con garantías y se da un canon de control "mucho más" estricto, e incluso excepcionalmente estricto. Se analizan supuestos complejos, como el uso por el sistema de IA público de "proxies" o datos afines a los especialmente sospechosos de discriminación, así como los tratos diferentes a grupos que los algoritmos generan dinámicamente. La prueba de la discriminación y sesgo algorítmico es un tema

ORCID: 0000-0003-2661-0010. Senior Research Fellow valgrAI: Valencian Graduate School and Research Network of Artificial Intelligence. El presente estudio es resultado de investigación del proyecto MICINN Retos "Derechos y garantías frente a las decisiones automatizadas... (RTI2018-097172-B-C21); grupo de investigación de excelencia Generalitat Valenciana "Algoritmic law" (Prometeo/2021/009, 2021-24); estancia (AEST/2021/012) y proyecto "Derecho, Cambio Climático y Big Data", Grupo de Investigación en Derecho Público y TIC como investigador de la Universidad Católica de Colombia; MININN "Transición digital de las Administraciones públicas e inteligencia artificia" (TED2021-132191B-I00) Y "Algorithmic Decisions and the Law: Opening the Black Box" (TED2021-131472A-I00) 2022-2023.

complejo. Inicialmente quien alega la discriminación debe aportar indicios. A partir de ahí, se sostiene que cuando el sector público usa IA hay una inversión general de la carga de la prueba; a partir de ahí se concreta lo que debe probar el proveedor o el usuario del sistema de IA público para que se entienda que no ha discriminado. También, se exponen las mejores prácticas contra la discriminación algorítmica y se afirma que muchos de estos remedios preventivos, así como los estudios de impacto, son ya obligatorios en el diseño de los usos públicos de la IA. Igualmente se recuerdan las futuras obligaciones del reglamento de datos y de IA de la UE frente a los sesgos. Si la prevención no es suficiente, se detalla qué sucede cuando se detecta una discriminación algorítmica pública: mitigación, corrección e incluso cese de uso del sistema de IA público; nulidad, anulabilidad e inderogabilidad de los sistemas de IA, los efectos de la discriminación algorítmica en el contexto de la contratación administrativa y algunas notas sobre la responsabilidad por el uso de IA público. Finalmente, se detalla quiénes pueden reaccionar ante una discriminación algorítmica y qué autoridades deben actuar.

Abstract: Some examples of biases and algorithmic discrimination, and the usual causes for their generation are presented. It is also set out the criteria for when public uses of AI can have an impact and relevance as discriminatory. The different concepts of discrimination are projected to the field of algorithms, underlining that it will generally be discrimination "by error". The cases in which so-called "indirect" discrimination can be useful are explained as well as "algorithmic positive actions" are explained, combined with differentiation actions and modulations and compensations to minimize discrimination. On these bases, the criteria for when a public algorithmic system is discriminatory are detailed. First, when it is NOT especially prohibited discrimination, it is asserted that the legal control overlaps especially with the control of arbitrariness and the requirements of motivation. This is particularly the case where AI systems are used for discretionary powers. Secondly, when a public algorithmic system DOES take into account particularly suspicious circumstances, it is supported that the guarantees generally overlap with the guarantees for specially protected data processing. Thus, a law with quality is imposed and a strict scrutiny, even exceptionally strict. Complex cases are analyzed, such as the use by the public AI system of "proxies" or related data, as well as different processing of dynamic algorithmic groups. Proof of algorithmic discrimination and bias is a complex issue. Initially, the party alleging discrimination must provide some clue. From there, there is a general reversal of the burden of proof when the public sector uses AI. It is specified what the administration must prove about the proper functioning of the public IA system. Best practices against algorithmic discrimination are set out; many of these preventive remedies are already mandatory in the design of public uses of AI, so are impact studies. Likewise, future EU data and AI regulation obligations against bias are recalled. If prevention is not enough, it is detailed what happens when public algorithmic discrimination is detected: mitigation, correction and even cessation of use of the public AI system; legal nullity of AI systems, the effects of algorithmic discrimination in the context of administrative procurement and some notes on liability for the use of public AI. Finally, there it is specified who can react to algorithmic discrimination and what authorities must act.

Palabras-clave: inteligencia artificial, algoritmos, administración, gobierno, igualdad, no discriminación, derechos fundamentales, sesgos

Key-words: artificial intelligence, algorithms, administration, government, equality, non-discrimination, fundamental rights, bias

SUMARIO.- 1.- SESGOS Y DISCRIMINACIÓN ALGORÍTMICA ¿DE QUÉ HABLA-MOS? Y ALGUNOS EJEMPLOS. 1.1.- Sesgos, errores y discriminación algorítmica y la conveniencia de abordarlos conjuntamente. 1.2.- Algunos ejemplos de discriminación algorítmica pública. 2.- ¿POR QUÉ SE GENERAN LA DISCRIMINACIÓN Y LOS SES-

GOS ALGORÍTMICOS?. 3.- ; RESPECTO DE OUÉ IA PÚBLICA ES EXIGIBLE LA NO DISCRIMINACIÓN?. 3.1.- La igualdad es exigible respecto de los sistemas de IA públicos, sus usos y sus efectos, aunque resulte difícil su categorización, 3.2.- El grado de legalidad exigible para el diseño, implantación y uso del sistema de IA público. 4.- ¿CUÁN-DO EL USO PÚBLICO DE UN SISTEMA DE IA TIENE IMPACTO Y SUFICIENTE RELEVANCIA JURÍDICA Y PUEDE CONSTITUIR UNA DISCRIMINACIÓN ALGO-RÍTMICA?. 4.1.- Los criterios y variables para determinar el impacto, nivel de riesgo v relevancia jurídica de los sistemas algorítmicos públicos. 4.2.-¿Cuándo el sesgo y el error algorítmico es relevante para constituir una discriminación? Datos estadísticos en discriminaciones indirectas y la necesidad de "recalibrarlos" a la baja para la IA pública. 5.- DISCRIMINACIONES ALGORÍTMICAS POR ERROR, DIRECTAS, INDIREC-TAS, INTERESECCIONALES Y ACCIONES POSITIVAS ALGORÍTMICAS. 5.1.- La discriminación o sesgo algoritmo será en la mayoría de supuestos "discriminación por error". 5.2.- Discriminaciones algorítmicas directas, interseccionales y por asociación. 5.3.- La complejidad de las discriminaciones algorítmicas indirectas, 5.4.- Acciones positivas algorítmicas, deber de diferenciación y modulaciones y compensaciones para la minimización de la discriminación. 6.- ¿CUÁNDO UN SISTEMA ALGORÍTMICO PÚ-BLICO QUE NO TIENE EN CUENTA CIRCUNSTANCIAS ESPECIALMENTE SOS-PECHOSAS ES DISCRIMINATORIO?. 6.1.- La igualdad "en" la ley o norma que habilite el uso del sistema de IA y el control laxo de la posibilidad de establecer diferencias de tratos, razonables y no arbitrarios. 6.2.- Control de discriminación y arbitrariedad y exigencias de motivación en los casos de aplicación algorítmica de la ley. Decisiones discrecionales y supuestos no especialmente sospechosos. 7.- ¿CUÁNDO UN SISTEMA ALGORÍTMICO PÚBLICO QUE SÍ TIENE EN CUENTA CIRCUNSTANCIAS ESPE-CIALMENTE SOSPECHOSAS ES DISCRIMINATORIO?. 7.1.- Las garantías frente a las discriminaciones algorítmicas especialmente prohibidas se superponen por lo general a las garantías de los tratamientos de datos especialmente protegidos. 7.2.- En la mayoría de casos será necesaria una ley de calidad y con garantías que habilite a que un sistema de IA pueda diferenciar en razón de datos especialmente sospechosos de discriminación. 7.3.- Un canon de control "mucho más" estricto, e incluso excepcionalmente estricto. 7.4.- La necesidad de regular legalmente el uso de datos especialmente protegidos para poder evaluar y minimizar la posible discriminación. 8.- ZONAS GRISES. GARAN-TÍAS ANTE EL USO POR EL SISTEMA DE IA PÚBLICO DE "PROXIES" Y GRUPOS ALGORÍTMICOS. 8.1.- Tratamientos diferentes a partir de "proxies" o datos afines o indirectos de grupos especialmente protegidos. 8.2.- Creación dinámica de grupos y colectivos humanos y la privacidad y no discriminación colectiva y de grupo. 9.-¿CÓMO PROBAR LA DISCRIMINACIÓN ALGORÍTMICA PÚBLICA? INVERSIÓN GENE-RAL DE LA CARGA DE LA PRUEBA Y LA NECESIDAD DE LA TRANSPARENCIA Y EXPLICABILIDAD, 9.1,- Los indicios que debe aportar la persona afectada por el sistema de IA público. Inversión general de la carga de la prueba y sus excepciones. 9.2.-Lo que debe probar el proveedor o el usuario del sistema de IA público. 10.- MÁS VA-LE PREVENIR QUE DISCRIMINAR. ¿CUÁLES SON LOS REMEDIOS Y MEJORES PRÁCTICAS PARA EVITAR LA DISCRIMINACIÓN ALGORÍTMICA PÚBLICA?. 11.-LA (CRECIENTE) OBLIGACIÓN DE DISEÑO Y ESTUDIOS DE IMPACTO FRENTE A LA DISCRIMINACIÓN ALGORÍTMICA DE LOS SISTEMAS DE IA PÚBLICOS. 11.1.- Los deberes de responsabilidad proactiva, privacidad en el diseño o estudio de impacto ya son exigibles en general a los usos públicos de la IA. 11.2.- Los usos públicos de IA de alto riesgo habrán de cumplir, además, los deberes del RIA. 11.3.- La extensión de estudios de impacto y la obligatoriedad de medidas para la no discriminación de los sistemas de IA públicos. 11.4.- La necesidad de una visión social y colectiva y de integrar garantías de calidad de datos y de participación. 12.- Y SI LOS REMEDIOS NO VALEN, ¿QUÉ PASA CUANDO SE DETECTA UNA DISCRIMINACIÓN ALGORÍTMICA PÚ-BLICA?. 12.1.- Obligaciones de mitigación, rectificación, corrección e incluso cese de uso

del sistema de IA público. 12.2.- Nulidad, anulabilidad e inderogabilidad de los sistemas de IA público y sus efectos. 12.3.- Los efectos de la discriminación algorítmica en el contexto de la contratación administrativa. 12.4.- Responsabilidad por el uso de IA público. 13.-¿QUIÉNES PUEDEN REACCIONAR ANTE UNA DISCRIMINACIÓN ALGORÍTMICA Y QUÉ AUTORIDADES DEBEN ACTUAR?. BIBLIOGRAFÍA CITADA.

## 1.- SESGOS Y DISCRIMINACIÓN ALGORÍTMICA ¿DE QUÉ HABLAMOS? Y ALGUNOS EJEMPLOS

#### 1.1.- Sesgos, errores y discriminación algorítmica y la conveniencia de abordarlos conjuntamente

Valga como definición introductoria de la "discriminación algorítmica" como la que "se produce cuando los sistemas automatizados contribuyen a un trato diferente injustificado" o, en su caso, un "impacto que desfavorece a las personas" en función de circunstancias específicas especialmente sospechosas de discriminación (raza, color, sexo, religión, edad, origen nacional, etc.) de modo contrario a Derecho².

El concepto de sesgo algorítmico se ha definido como "una anomalía en la salida de los sistemas de IA, debido a los prejuicios o suposiciones erróneas realizadas durante el proceso de desarrollo del sistema o prejuicios en los datos de entrenamiento, por lo que los resultados del sistema de IA no pueden generalizarse ampliamente."<sup>3</sup> Para ISO es "el grado en que un valor de referencia se desvía de la

WHITE HOUSE (2022): The Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People, White House Office of Science and Technology Policy, Whasington DC., octubre, pp. 10 y ss. https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdfl

<sup>&</sup>lt;sup>3</sup> JOINT RESEARCH CENTRE; ESTÉVEZ ALMENZAR, M. y otros (2022): *Glossary of human-centric artificial intelligence*, Joint Research Centre, Unión Europea, Luxemburgo, pp. 16-17 https://publications.jrc.ec.europa.eu/repository/handle/JRC129614 DOI:10.2760/860665

verdad"<sup>4</sup>. EL Grupo Independiente de Expertos de Alto nivel sobre Inteligencia Artificial (HLEG) señala que "Un sesgo es una inclinación que favorece o perjudica a una persona, objeto o posición."<sup>5</sup>

En una de las mejores aportaciones sobre el tema, el NIST aprecia más de cuarenta variedades o tipologías de sesgos<sup>6</sup> y los agrupa en tres categorías dominantes: los sesgos sistémicos, los sesgos estadísticos y computacionales y los sesgos humanos, a saber.

- Los sesgos sistémicos son esencialmente sesgos institucionales o históricos. No son voluntarios. Aunque se sigan las reglas o normas existentes. Están en los conjuntos de datos y en las normas, prácticas y procesos institucionales a lo largo del ciclo de vida de la IA y en la cultura y la sociedad en general.
- Los sesgos estadísticos y computacionales se dan cuando la muestra no es representativa de la población. El error puede deberse a datos heterogéneos, representación de datos complejos en representaciones matemáticas más simples, datos erróneos y sesgos algorítmicos como el ajuste excesivo o insuficiente, el

ISO (2006): Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability, ISO, Tech. Rep. ISO 3534-1:2006, https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/04/01/40145.html

<sup>5</sup> HLEG (Comisión Europea), (2019): Directrices éticas para una IA fiable, Comisión Europea 2019, p. 48 https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

NIST; SCHWARTZ, R. y otros (2022): Towards a Standard for Identifying and Managing Bias in Artificial Intelligence, Special Publication 1270 (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, pp. 6-9 https:// doi.org/10.6028/NIST.SP.1270

Así en su amplio glosario (pp. 49-55): de actividad, de comportamiento, cognitivo, de amplificación, de anclaje, de información, de automatización, de disponibilidad, de confirmación, del consumidor, de producción de contenidos, de dragado de datos, de generación de datos, de despliegue, de detección, emergente, de evaluación, de exclusión, de bucle, de financiación, histórico, de información humana, implícito, heredado, institucional, de interpretación, en el procesamiento, de enlace, de pérdida de conciencia de la situación, de medición, de confusión de modo, de selección del modelo, de popularidad, de población, de presentación, de clasificación, de representación, social, de supervivencia, temporal, de incertidumbre, de interacción del usuario.

tratamiento de valores atípicos y la limpieza de datos y los factores de imputación.

- Los sesgos humanos, errores sistemáticos en el pensamiento humano. Son una parte fundamental de la mente humana, así se dan sesgos de anclaje, de disponibilidad, el sesgo de confirmación, entre muchos otros. Asimismo se trata de atajos mentales adaptativos. Obviamente, no son exclusivos de las interacciones humanas con la IA, pero en este ámbito, se insertan en todo el ciclo de vida de la IA y en su uso.

Considero que procede abordar los sesgos y errores juntamente con la discriminación algorítmica. Cualquier uso de sistemas de IA habitualmente genera errores y sesgos, en algunos casos estos pueden tener impactos relevantes. Todo sesgo de cierta relevancia en el uso de la IA potencialmente puede suponer un trato diferente no justificado. Y como se verá, ello puede constituir una discriminación por error, una discriminación especialmente prohibida con relación a circunstancias especialmente sospechosas o una discriminación general.

#### 1.2.- Algunos ejemplos de discriminación algorítmica pública

Los supuestos de discriminación algorítmica son de lo más variado<sup>7</sup>. En contextos más relacionados con el uso público de IA, con los sistemas de identificación biométrica e IA, las posibilidades de

Del ámbito público y privado, Se pueden seguir diversos ejemplos en CONSEJO DE EUROPA; ZUIDERVEEN BORGESIUS F (2018): Discrimination, artificial intelligence, and algorithmic decision-making. Directorate General of Democracy, CONSEJO DE EUROPA, pp. 23-31, https://rm.coe. int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73

BID- Universidad Adolfo Ibáñez (2022): Formulación ética de proyectos de ciencia de datos, Banco Interamericano de Desarrollo, División de Gobierno Digital y Universidad Adolfo Ibáñez, pp. 6-9, https://goblab.uai.cl/guia-formulacionetica-de-proyectos-de-ciencia-de-datos/

WHITE HOUSE (2022), op.cit., pp. 23-24 y pp. 29-30.

EXECUTIVE OFFICE OF THE PRESIDENT (2016): Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights. May 2016. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016\_0504\_data\_discrimination.pdf

O'NEIL, C. (2017): Weapons of Math Destruction. Penguin Books. 2017.

generar errores, sesgos y discriminaciones son muchas, aunque sean involuntarias. Estos sistemas son probabilísticos y aplicados a miles o millones de personas los errores pueden salir muy caros a los afectados. En Buenos Aires Guillermo Federico Ibarrola fue erróneamente identificado por un sistema inteligente como prófugo y estuvo detenido 6 días; el 12 de abril de 2022 un juez suspendió el sistema que fue anulado por sentencia de 7 de septiembre de 2022. La misma suerte y los mismos días que estuvo detenido Robert Williams en Detroit por otro falso positivo de reconocimiento facial8. Escáneres corporales de aeropuertos en EEUU (TSA) han marcado a los viajeros transgénero como más probables sospechosos para efectuar un control particular.9. Servicios de empleo también en EEUU han generado la discriminación de discapacitados<sup>10</sup>. En el ámbito de la salud, un sistema aplicado aproximadamente 200 millones de estadounidenses cada año preferenciaba para atención sanitaria temprana a quienes habían costado más dinero anteriormente. Como las personas negras recibían

Ya en 2008, en el sector público pueden seguirse supuestos reales en CITRON, D. K. (2008): "Technological Due Process", 85 Wash. U. L. Rev. pp. 1249-1313, pp. 1256, http://openscholarship.wustl.edu/law\_lawreview/vol85/iss6/2

Al respecto pueden seguirse mis estudios recientes sobre el tema con remisión a toda la literatura "Sistemas de inteligencia artificial con reconocimiento facial y datos biométricos. Mejor regular bien que prohibir mal", en *El Cronista del Estado Social*, IUSTEL, monográfico Inteligencia artificial, nº 100, septiembreoctubre 2022, pp. 68-79. Una versión extensa, en la obra que coordino con Francisco Balaguer.

Al respecto, a partir de los supuestos señalados en WHITE HOUSE (2022), op.cit., cabe seguir, CHRISTIAN, A. (2022): Four Ways the TSA Is Making Flying Easier for Transgender People. American Civil Liberties Union. https://www.aclu.org/news/lgbtq-rights/four-ways-the-tsa-is-making-flying-easier-for-transgender-people

U.S. TRANSPORTATION SECURITY ADMINISTRATION (2022): *Transgender/ Non Binary / Gender Nonconforming Passengers*. TSA, https://www.tsa.gov/transgender-passengers

U.S. EQUAL EMPLOYMENT OPPORTUNITY COMMISSION (2022): The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees. EEOC-NVTA-2022-2, https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence

U.S. DEPARTMENT OF JUSTICE (2022): Algorithms, Artificial Intelligence, and Disability Discrimination in Hiring. Mayo, https://beta.ada.gov/resources/ai-guidance/

menos asistencia médica y habían hecho menos gasto, eran postergados por el sistema respecto de los blancos con enfermedades y necesidades similares<sup>11</sup>. Son ya conocidos los supuestos de sesgos raciales en el ámbito policial y de aplicación de la ley12. Así, la predicción de reincidencia de los individuos bajo custodia federal sobreestimaba o subestimaba a diversos grupos de color y hubo de corregirse<sup>13</sup>. No menos del 70% de estados EEUU utilizan desde 2004 algún instrumento mecánico para la decisión de libertad condicional (ej.: Level of Services Inventory-Revised, LSI-R). Los sistemas de predicción de riesgo (risk assessment instruments, Public Safety Assessment, PSA) después de la condena se emplean en más de veinte jurisdicciones de EEUU en los últimos años. Algunos de estos sistemas se usan por jueces para determinar la condena por el riesgo de reincidencia, siendo especialmente conocido el sistema COMPAS "Correctional Offender Management Profiling for Alternative Sanctions". En el mismo, los negros tienen casi el doble de probabilidades que los blancos de ser

OBERMEYER, Z. y otros (2019), Dissecting racial bias in an algorithm used to manage the health of populations. Science. Vol. 366, No. 6464. Oct. 25, 2019 https://www.science.org/doi/10.1126/science.aax2342

Sobre el uso penal, policial y de aplicación de la ley, la literatura es muy abundante, me permito destacar: MARTÍNEZ GARAY, L. y MONTES SUAY, F. (2018): "El uso de valoraciones del riesgo de violencia en Derecho Penal:: algunas cautelas necesarias", *Indret: Revista para el Análisis del Derecho*, n°. 2; BARONA, S. (2021): *Algoritmización del derecho y de la justicia: De la inteligencia artificial a la Smart Justice*, Tirant lo Blanch, Valencia; SIMÓ, P. (2022): *La prisión algorítmica*, Tirant lo Blanch, Valencia; SIMÓ, P. y MAGRO, V. (2021): *Justicia cautelar e inteligencia artificial: la alternativa a los atávicos heurísticos judiciales*, J.M. Bosch Editor, Barcelona, 2021; MIRÓ LLINARES, F. (2022): "Inteligencia artificial, delito y control penal: nuevas reflexiones y algunas predicciones sobre su impacto en el derecho y la justicia penal", en *El Cronista del Estado Social y Democrático de Derecho*, n°. 100, 2022, pp. 174-183.

JOHNSON, C. (2022 a): Flaws plague a tool meant to help low-risk federal prisoners win early release. NPR. 26 enero, 2022 https://www.npr. org/2022/01/26/1075509175/flaws-plague-a-tool-meant-to-help-low-risk-federal-prisoners-win-early-release y (2022 b): Justice Department works to curb racial bias in deciding who's released from prison. NPR. 19 de abril, https://www.npr.org/2022/04/19/1093538706/justice-department-works-to-curb-racial-bias-in-deciding- whos-released-from-pris;

NATIONAL INSTITUTE OF JUSTICE (2021): Review and Revalidation of the First Step Act Risk Assessment Tool. National Institute of Justice NCJ 303859. https://www.ojp.gov/pdffiles1/nij/303859.pdf

etiquetados como de mayor riesgo de reincidir. Entre los blancos se produce el error contrario. El uso de este sistema dio lugar al famoso caso *Loomis*<sup>14</sup>.

En otros ámbitos del uso público de IA, ya hace bastantes años se mencionaban errores masivos por sistemas automatizados en la asignación de ayudas públicas (como el sistema *CalWIN* en California o en Texas o el CBMS de Colorado (*Colorado Benefits Management System*). Unos 1.500 pasajeros semanalmente no podían subir a un avión al ser incluidos erróneamente en la lista "No Fly", o excluidos de contratación pública<sup>15</sup>. En EEUU, la sentencia *K.W. V. Armstrong* 89 F.3D 962, 976 (9TH Cir. 2015) revisó el sistema del *Department of Health and Welfare* que atribuía beneficios de seguridad social por enfermedad en Idaho y se detectaron diferencias desproporcionadas entre distintas zonas y errores estadísticos en la fórmula utilizada.

En el ámbito educativo también es habitual el uso de sistemas IA. Y con ellos los sesgos y la discriminación. Así, un sistema de predicción de abandono de estudios utilizado en más de 500 universidades de EEUU generaba sesgos que multiplicaban por cuatro la predicción respecto de las personas de color¹6. En el ámbito de empleo, los sistemas que predicen quién tiene más probabilidades de hacer clic en un anuncio de trabajo dirigían a las mujeres a trabajos de cajeras de supermercado y a los negros a empresas de taxis.¹7

Sentencia State v. Loomis, 881, N.W.2d 749, 7532 (Wis, 2016).
Por todos, DE MIGUEL BERIAIN, I. (2018). "Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling, *Law, Probability and Risk, Volume 17, Issue 1, 1* March 2018, pp. 45–53, https://doi.org/10.1093/lpr/mgy001 Ver también, CONSEJO DE EUROPA (2018), *op.cit.*, pp. 23-24. Sobre la herramienta, http://www.equivant.com

<sup>&</sup>lt;sup>15</sup> CITRON, D. K. (2008), *ob.cit.*, p. 1256.

FEATHERS, T. (2021): "Major Universities Are Using Race as a "High Impact Predictor" of Student Success: Students, professors, and education experts worry that that's pushing Black students in particular out of math and science", en *The Markup*. Marzo https://themarkup.org/machine-learning/2021/03/02/major-universities-are-using-race-as-a-high-impact-predictor-of-student-success

BOGEN, M. (2019): "All the Ways Hiring Algorithms Can Introduce Bias". Harvard Business Review. Mayo. https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias

Siempre es más fácil ver la paja en el ojo ajeno que en el propio. Y los casos conocidos a este lado del charco son muchos. Así, en Francia el sistema de admisión Parcoursup distribuye estudiantes en establecimientos de educación superior con base en criterios previstos legalmente. En 2020 el Consejo Constitucional francés reconoció el acceso a las características del sistema por cualquier persona y que cada institución de educación superior debe informar sobre los criterios según los cuales se han examinado las solicitudes de *Parcoursup*. 18 En 2020 en Reino Unido por el Covid no se realizaron los exámenes A-Level de acceso a la universidad. El organismo regulador (Ofqual) creó un sistema basado en un algoritmo que afectó a casi un millón de personas. Los problemas que se generaron llevaron a la dimisión del Ministro de Educación. La crítica más habitual es que se penalizó a los alumnos de colegios públicos. El diseño del algoritmo dio lugar a "sesgos técnicos que no se identificaron ni se corrigieron de forma proactiva"19.

Es también famoso el sistema de IA público de empleo austríaco PAMAS. El sistema -correctamente- detectaba que ser mujer y de color eran factores de baja probabilidad de encontrar trabajo. El problema fue humano, dado que se programó el sistema para dedicar menos recursos a estos supuestos de baja probabilidad, de este modo se perpetuaba la exclusión social<sup>20</sup>. De modo bastante similar, el sis-

Criterios regulados en artículo L. 612.3 del Código de educación. Sobre el tema https://es.frwiki.wiki/wiki/Parcoursup. La Decisión del Consejo Constitucional https://www.conseil-constitutionnel.fr/actualites/communique/decision-n-2020-834-qpc-du-3-avril-2020-communique-de-presse

Analiza el caso exhaustivamente GUTIÉRREZ DAVID, M.E., "Acceso al código fuente y a los algoritmos de las Administraciones inteligentes. Lecciones a partir de experiencias comparadas", en COTINO HUESO L. y CASTELLANOS CLARAMUNT J. (coords.), *Transparencia y explicabilidad de la inteligencia artificial*, Tirant lo Blanch, Valencia, 2022, pp.135-167. La cita es de WACHTER, S., y otros, "Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law", *West Virginia Law Review*, Vol. 123, núm. 735, 2021, p. 739. https://x.gd/hKKKqS.

Sobre el tema, SORIANO ARNANZ, A. (2022 a): "El uso de sistemas automatizados por los poderes públicos. El caso del sistema empleado por el servicio público de empleo austriaco", en COTINO HUESO, L. y TODOLÍ SIGNES, A. (coord.), Explotación y regulación del uso del big data e inteligencia artificial para los servicios públicos y la ciudad inteligente, Tirant lo Blanch, Valencia, pp. 192-216.

tema polaco de empleo introdujo en 2014 un sistema con perfilados (*Publiczne Służby Zatrudnienia*, PSZ). Un algoritmo decidía la categoría y puntuación final que otorgaba al desempleado y a partir de tal puntuación el personal del servicio había de tomar decisiones. Sin embargo, el sesgo de automatización llevó a que habitualmente los funcionarios siguieran estos perfiles y puntuaciones automáticas sin mayor supervisión. El sistema no era transparente de cara al público. La Oficina Superior de Auditoría (*Najwyższa Izba Kontroli*) que lo supervisó detectó la ineficacia del sistema y que las mujeres se puntuaban diferente a los hombres, en general, las personas más vulnerables se atribuían al grupo con más dificultades de recibir asistencia del sistema<sup>21</sup>.

El sistema neerlandés SyRI fue especialmente conocido por generar la decisión judicial de uso público de sistemas algorítmicos más importante en Europa<sup>22</sup>. El Systeem Risicoindicatie (SyRI) es un sistema automatizado para detectar diversas formas de fraude, incluidos beneficios sociales, subsidios y fraude fiscal. El sistema podía llegar a ser de "autoaprendizaje y aprendizaje profundo", "procesaba una -casi ilimitada- cantidad de datos sustancial" (6.50) de muchas categorías al punto que "es difícil imaginar cualquier tipo de datos personales que no sean elegibles para el procesamiento en SyRI" (6.98). Y para generar las alertas de posibles ilegalidades no había "intervención humana significativa", luego ya la alerta era investigada para en su caso generar consecuencias administrativas. SyRI contaba con amplia regulación legal y reglamentaria y, todo sea dicho, con muchas garantías, incomparablemente superiores a las que puede haber en España, tanto por los desconocidos sistemas de TGSS y AEAT como los más conocidos como el valenciano SALER<sup>23</sup>. Sin embargo, se consideró

https://algorithmwatch.org/en/poland-government-to-scrap-controversialunemployment-scoring-system/

Sentencia de 5 de febrero de 2020 del Tribunal de Distrito de la Haya (C / 09/550982 / HA ZA 18-388). Me remito a mi estudio (2020): "'SyRI, ¿a quién sanciono?' Garantías frente al uso de inteligencia artificial y decisiones automatizadas en el sector público y la sentencia holandesa de febrero de 2020, en La Ley Privacidad, Wolters Kluwer nº 4, mayo. Acceso en Academia

Con bastante profundidad, mi estudio (2020 a) "Hacia la transparencia 4.0: el uso de la inteligencia artificial y big data para la lucha contra el fraude y la corrupción y las (muchas) exigencias constitucionales", en Carles Ramió (coord.),

que su falta de transparencia generaba vulneraciones de la igualdad, la privacidad y el debido proceso. En el proceso participó el Relator Especial de la ONU precisamente por sus efectos discriminatorios y consideró que SyRI "tiene un efecto discriminatorio y estigmatizador", pues al investigar vecindarios más marginados "contribuye a los estereotipos y refuerza una imagen negativa de los ocupantes de dichos vecindarios". La sentencia afirma que el sistema era discriminatorio por falta de transparencia (6.92), "la transparencia, en aras de la verificabilidad» es necesaria frente a los peligros de «exclusión o discriminación injustificadas» (6.91).

En España, entre otros supuestos, el Gobierno estatal creó la aplicación BOSCO, cedida a las compañías eléctricas para verificar los requisitos de los solicitantes para la atribución del bono social eléctrico. Su funcionamiento era más que sospechoso de errores, que no se han podido probar plenamente por la reiterada falta de transparencia sobre el mismo<sup>24</sup>. También en nuestro país, el sistema *VeriPol* estima la probabilidad de que una denuncia sea falsa y el sistema *VioGén* señala la peligrosidad de posibles hombres maltratadores y puede acabar determinando las decisiones de prisión preventiva, así como las decisiones de política penitenciaria en Cataluña el sistema *RisCanvi*. La mayor o menor opacidad de estos sistemas no permite valorar ni su incidencia efectiva en las decisiones que se adoptan ni la posible existencia de sesgos o errores que puedan generar discriminación<sup>25</sup>.

En el ámbito privado son bien conocidos los sistemas de IA que han generado sesgos y discriminación en el ámbito de la contra-

Repensando la administración digital y la innovación pública, Instituto Nacional de Administración Pública (INAP), Madrid, 2021. https://links.uv.es/FUW2pz6

Me remito a los diversos estudios en COTINO HUESO L. y CASTELLANOS CLARAMUNT J. (coords.) (2022), ob. cit., en particular GUTIÉRREZ DAVID, M.E.,), ob. cit., pp. 135-167 y VESTRI, G. (2022): "Denegación vs. Derecho de acceso al código fuente en los sistemas algorítmicos. Una perspectiva jurídico-administrativa", pp. 111-134. Y (2022): "El acceso a la información algorítmica a partir del caso bono social vs. Fundación Ciudadana Civio", Revista General de Derecho Administrativo, nº 61, 2022.

Al respecto, me remito a la investigación coordinada por el equipo liderado por Lucía Martínez Garay Universidad de Valencia-Amnistía Internacional sobre las insuficiencias de estos sistemas.

tación, los buscadores, asociación de textos, generación de contenidos, etiquetado de imágenes, personalización de precios, de publicidad, análisis biométrico de sentimientos y lecturas de emociones, concesión de préstamos, fijación de primas de seguro y un largo etcétera. Baste ahora señalar que en muy buena medida estos sesgos y discriminaciones son claramente exportables a los usos públicos de IA.

### 2.- ¿POR QUÉ SE GENERAN LA DISCRIMINACIÓN Y LOS SESGOS ALGORÍTMICOS?

Los sesgos y errores se generan antes de concebirse un sistema de IA, en cualquiera de sus fases o bien en el uso posterior de estos sistemas por los usuarios o de sus resultados de salida<sup>26</sup>. Los motivos por los que un sistema de IA genera sesgos y discriminaciones pueden ser de lo más variado<sup>27</sup>. No obstante, como señalé hace años,

<sup>&</sup>lt;sup>26</sup> NIST; SCHWARTZ, R. y otros (2022), *op.cit.*, pp. 6-7.

Sobre las causas, además de la referida guía del NIST, resultan de especial interés los clásicos de O'NEIL, C. (2017), ob. cit., pp. 15-31; BAROCAS, S. y SELBST, A. D. (2016): "Big Data's disparate impact". 104 Calif Law Rev 671, http://dx.doi. org/10.2139/ssrn.2477899. Una buena descripción, TISCHBIREK, A. (2020): "Artificial Intelligence and Discrimination: Discriminating Against Discriminatory Systems", en WISCHMEYER, T y RADEMACHER. T. (ed.), (2020): Regulating Artificial Intelligence, Springer, pp. 103-121, ver pp. 104-109. https://doi. org/10.1007/978-3-030-32361-5; HLEG (2019), op.cit., p. 48, CONSEJO DE EUROPA (2018), op.cit., p. 15, nota. 28;

VALLE ESCOLANO, R., "Transparencia en la inteligencia artificial y en el uso de algoritmos: una visión de género", COTINO HUESO L. y CASTELLANOS CLARAMUNT J. (coords.) (2022), *ob. cit.*, pp. 85-110, ver pp. 96 y ss., 98-99; ZLOTNIK, A. (2019): "Inteligencia Artificial en las Administraciones Públicas: definiciones, evaluación de viabilidad de proyectos y áreas de aplicación", *Boletic*, núm. 84, pp. 27-28, http://apiscam.blogspot.com/2019/11/boletic-revista-de-astic-monografico-de.html así como los diversos trabajos de SORIANO ARNANZ, A. (2021 b): "Decisiones automatizadas: problemas y soluciones jurídicas. Más allá de la protección de datos", *Revista de Derecho Público: Teoría y Método*, Vol. 3, pp. 85-127, DOI:10.37417/RPD/vol\_1\_2021\_535 o SORIANO ARNANZ, A. (2022 b): "Discriminación algorítmica: garantías y protección jurídica", en COTINO HUESO, L. (coord.), *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*, Aranzadi, Cizur Menor, 2022, págs. 139-169.

los grandes problemas de la inteligencia artificial por lo general se deben a la *estupidez humana*<sup>28</sup>. Son los humanos quienes deciden crear el sistema, para qué usos, a partir de qué datos alimentarlo, de dónde obtenerlos, qué valor otorgarles a los datos y a los resultados que arroja el sistema.

Un problema importante es la "tendencia a confiar automáticamente o en exceso en la información de salida generada por un sistema de IA de alto riesgo ("sesgo de automatización", art. 13.4° b) RIA. El ser humano está sesgado de fábrica y ello se refleja en la forma en que el sistema fue programado o los datos seleccionados o procesados. Y especialmente se pueden dar los problemas y sesgos respecto de diversos colectivos si tres de cada cuatro profesionales, normalmente del ámbito STEM, son hombres blancos heterosexuales acomodados católicos-cristianos (77%). Sólo el 15 % de los científicos de datos son mujeres y menos del 10% entre los de científicos de datos que luego beben las máquinas) según ella misma proceden de hombre blanco, técnico, con educación formal, de habla inglesa, de 15 a 49 años, de un país cristiano y del hemisferio norte. Son muy intensas las infrarrepresentaciones.<sup>30</sup>

Una IA diseñada por colectivos de este tipo involuntariamente ha de tener falta de empatía hacia los colectivos tradicionalmente discriminados, así como "puntos ciegos", como los que tenemos con el retrovisor del vehículo.<sup>31</sup>

También, la causa de sesgos y discriminaciones puede ser un erróneo diseño y elección de algoritmos, del peso que atribuyen a unos u otros factores o a los errores en el desarrollo de los sistemas de apren-

COTINO HUESO, L. (2017): "Big data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales", en *Dilemata. Revista Internacional de Éticas Aplicadas*, nº 24, 2017. pp. 131-150, https://goo.gl/iERVha

<sup>29</sup> https://www.burtchworks.com/wp-content/uploads/2019/06/Burtch-Works-Study\_DS-PAP-2019.pdf

https://en.wikipedia.org/wiki/Wikipedia:Systemic\_bias#The\_%22average\_Wikipedian%22

Sobre este punto ciego, un clásico, PRONIN, E. y otros, (2002).: "The Bias Blind Spot: Perceptions of Bias in Self Versus Others", *Volume 28, Issue 3, SAGE*, https://doi.org/10.1177/0146167202286008

dizaje automático. Pese a sistemas y personas del todo adecuadas, es extremadamente difícil operacionalizar todos los aspectos relevantes de una realidad social. De ahí que los programadores pueden obviar algunas variables, introducir sesgos en el sistema o escoger bases de datos que ofrecen una representación parcial de la realidad. Frente a ello, los seres humanos tienen la capacidad de apreciar determinados elementos intangibles que no son captados por los sistemas automatizados.<sup>32</sup> Un sistema de IA correcto utilizado en un contexto diferente o para una población diferente para los cuales fue originalmente desarrollado puede generar problemas<sup>33</sup>.

Cabe recordar que la existencia de correlaciones que fácilmente encuentran los sistemas de *machine learning* no implica causalidad, por lo que deben revisarse las reglas que se generen. Por ejemplo, se suele mencionar con ánimo jocoso que hay una correlación casi perfecta entre el consumo de *mozzarela* en Estados Unidos y el número de doctorados en Ingeniería Civil entre 2000 y 2009, o la edad de Miss América y los asesinatos con objetos de calor o vapor<sup>34</sup>.

Con especial significación, los sesgos y discriminaciones se generan por causa de los datos de entrenamiento. Se afirma así "Garbage in, garbage out" (GIGO), "basura entra, basura sale"<sup>35</sup>: datos de baja calidad llevan a resultados baja calidad. El Parlamento UE en su resolución sobre macrodatos ha insistido en que "los datos de capacitación a menudo son de una calidad cuestionable y no son neutrales" (Letra B), la "baja calidad" de los datos o los procedimientos "podrían dar lugar a algoritmos sesgados, correlaciones falsas, errores, una subestimación de las repercusiones éticas, sociales y legales" (Considerando

En esta dirección, SORIANO ARNANZ, A. (2021 b), ob.cit., p. 90.

<sup>&</sup>lt;sup>33</sup> VALLE ESCOLANO, R. (2022), *ob. cit.*, p. 99.

http://tylervigen.com/spurious-correlations

https://es.wikipedia.org/wiki/Garbage\_in,\_garbage\_out. Ver FRA, (2018): UE. (2018). #BigData: Discrimination in data-supported decision making, European Union Agency for Fundamental Rights, Viena, https://fra.europa.eu/en/publication/2018/big-data-discrimination; FRA- Agencia de la Unión Europea para los Derechos Fundamentales, (2019): Data quality and artificial intelligence-mitigating bias and error to protect fundamental rights, Luxembourg, Publications Office, junio, https://fra.europa.eu/en/publication/2019/artificial-intelligence-data-quality

NIST; SCHWARTZ, R. y otros (2022), op.cit., pp. 15-16 y 17-20.

m)<sup>36</sup>. Muchos conjuntos de datos no son generalizables o representativos, sin embargo se utilizan para entrenar los sistemas. Los grupos desfavorecidos están sistemáticamente subrepresentados, así como los datos utilizados para el procesamiento del lenguaje natural difieren significativamente de sus aplicaciones del mundo real. También se alerta de que se reutilizan conjuntos de datos, a partir de sistemas ya diseñados con deficiencias o los generados por usuarios finales (por ejemplo, de redes sociales), por lo que el sistema es vulnerable a sesgos humanos o estadísticos aditivos. Los daños pueden perpetuarse y amplificarse. Las mismas series históricas de datos que alimentan al sistema de IA incorporan, reproducen y perpetúan estructuras de discriminación. Lo peor, además, es que se pueden generar espirales de sesgo, error y discriminación<sup>37</sup> pues los sistemas muy posiblemente acentuarán sus decisiones al nutrirse de nuevos datos cada vez más negativos para los sectores perjudicados.

Y, por supuesto, los sesgos y discriminaciones pueden provenir porque el proveedor del sistema de IA público sea privado y puede perder de objetivos de interés general como la protección de los segmentos más vulnerables.

## 3.- ¿RESPECTO DE QUÉ IA PÚBLICA ES EXIGIBLE LA NO DISCRIMINACIÓN?

## 3.1.- La igualdad es exigible respecto de los sistemas de IA públicos, sus usos y sus efectos, aunque resulte difícil su categorización

El derecho a la igualdad tiene efecto directo respecto de la actuación de los poderes públicos (art. 53.1 CE) y como se verá prohíbe la

PARLAMENTO EUROPEO (2017): Resolución de 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley (2016/2225(INI)). Acceso en http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0076+0+DOC+XML+V0//ES

<sup>37</sup> En particular BAROCAS, S. y SELBST, A. D. (2016), ob. cit. y los diversos trabajos de Soriano.

discriminación tanto en general cuanto la relativa a colectivos sospechosos de ser discriminados. La Ley 15/2022, de 12 de julio integral para la igualdad de trato y la no discriminación en principio sólo sería aplicable para los supuestos de discriminación y sesgos algorítmicos vinculados con las categorías especialmente sospechosas de discriminación pero no para los supuestos generales. En todo caso, seguir esta ley facilita la comprensión de la atracción para el ámbito jurídico del uso de sistemas públicos de IA. Sus obligaciones son de "aplicación al sector público" (art. 2.4° Ley 15/2022). Cualquier uso material de IA por el sector público quedaría expresamente en su ámbito de aplicación y obligaciones (art. 3), pues alcanzan expresamente a la "inteligencia artificial y gestión masiva de datos, así como otras esferas de análoga significación", además de "n) internet, redes sociales y aplicaciones móviles". De igual modo, yendo más allá del estricto procedimiento administrativo, la no discriminación también se exige en las actividades relativas al empleo público (b), educación (d), sanidad (e), seguridad ciudadana (h), administración de justicia (i), protección social, las prestaciones y los servicios (j) sociales, acceso, oferta y suministro de bienes y servicios a disposición del público (k), transporte (f) o cultura (g).

Más allá de la estricta actuación administrativa y procedimental, un "incidente discriminatorio" (art. 25. 3° Ley 15/2022) puede darse por "toda disposición, conducta, acto, criterio o práctica" (art. 6 Ley 15/2022), así como "la inacción, dejación de funciones, o incumplimiento de deberes" (art. 4.1° Ley 15/2022).

Como más adelante se señala, la existencia de un sistema de IA público podrá y en ocasiones deberá estar regulado por leyes o reglamentos y disposiciones generales. En algunos casos estas normas podrán ser objeto de posible impugnación directa o indirecta, especialmente si las normas regulan la posibilidad de establecer diferencias de trato. En la mayoría de los casos la discriminación o sesgos algorítmicos se deberán a decisiones adoptadas o no en el desarrollo, implantación y uso del sistema probablemente no formalizadas y de más que de dudosa naturaleza jurídica. O en su caso, por el funcionamiento autónomo del sistema de IA. En cualquier caso, serán "criterios o prácticas" que pueden ser discriminatorios. Aunque la existencia del sistema de IA o las decisiones sobre el su funcionamiento o efectos no adquieran forma

jurídica o formalmente no formen parte propiamente del procedimiento, ello no es obstáculo para generar una discriminación.

No es necesario que una discriminación o sesgo algorítmico se vincule necesariamente a actuaciones administrativas "en el marco de un procedimiento administrativo" como reza el artículo 41 LRISP. Ahora bien, si se está en el terreno del artículo 41 LRJSP se habrá determinado el órgano competente del sistema "que debe ser considerado responsable a efectos de impugnación" (art. 41.2° LR JSP). También en el ámbito de este precepto, en el caso de la AGE puede focalizarse la atención en la "resolución" de autorización del sistema (art. 13. 2º Real Decreto 203/2021, RAFESP), igual que en el caso de Cataluña (art. 54.2° Decreto 76/2020, de 4 de agosto, de Administración digital de Cataluña). En el ámbito de la contratación de sistemas de IA, hay que acudir a las vías de impugnación que pudieran darse en este contexto, así como a la posibilidad de exigencia de responsabilidades a los contratistas. Como señala Boix, no está muy claro si se puede impugnar el algoritmo o programa empleado en sí mismo, o solo es posible atacar sus actos de aplicación<sup>38</sup>. El uso del sistema de IA en su caso puede ubicarse en las diligencias preliminares, en muchas ocasiones puede ser considerado un acto de trámite como un eslabón del procedimiento. En todo caso, de cara a su impugnabilidad habría que determinar la fase de procedimiento específica, su naturaleza como acto de iniciación, ordenación, o de instrucción y distinguir si tiene mayor o menor sustantividad propia. Habría que determinar el tipo de acto de trámite, si es acto de terminación o firme. En cualquier caso, ante dificultades de esta categorización siempre se podrá acudir a los conceptos habituales y menos formales en el ámbito de la no discriminación.

Igualmente, la discriminación se puede dar por "inacción" (art. 4.1° Ley 15/2022) o el "incumplimiento de los deberes", como puedan ser las obligaciones de responsabilidad proactiva en el diseño, control o evaluación del sistema de IA que exige la normativa de protección de datos u otras regulaciones que luego se analizan. La

BOIX PALOP, A. (2020): "Los algoritmos son reglamentos: la necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones", en Revista de Derecho Público: Teoría y Método, Vol. 1, 2020, p. 261 https://doi.org/10.37417/RPD/vol\_1\_2020\_33

inacción puede ser atacable por las vías de acción frente a la inactividad de la Administración (art. 29 LJCA). En tanto en cuanto está implicado el artículo 14 CE en principio será una posibilidad acudir directamente a la jurisdicción contenciosa a través del procedimiento para la protección de los derechos fundamentales (arts. 114 y ss.). Y no genera particular problema acudir a la vía de reclamación de responsabilidad patrimonial por el funcionamiento normal o anormal de los servicios públicos en razón del uso de sistemas de IA (arts. 32 y ss. LRJSP). En este punto también es de interés recordar que "acreditada la discriminación se presumirá la existencia de daño moral" (art. 27 Ley 15/2022, de 12 de julio) y cabe acudir al régimen de la responsabilidad patrimonial de las Administraciones Públicas.

En muchos supuestos, una vía de *aterrizar* jurídicamente el uso de un sistema de IA público será el tratamiento de datos personales que ello implica. Esencialmente: cuando la salida o decisión del sistema de IA público se individualice y aplique a una persona física; así como cuando el sistema de IA público suponga la elaboración de perfiles o decisiones automatizadas a partir de datos personales facilitados, o también los inferidos por el sistema. Igualmente habrá tratamiento de datos en la mayoría de los supuestos de personalización y servicios 360° de la Administración y un largo etcétera. En estos supuestos y como consecuencia, ello supondrá la aplicación en bloque del régimen de protección de datos al sistema de IA público. Y como se verá, una discriminación algorítmica muy posiblemente implicará el incumplimiento de los principios básicos como el de calidad de los datos o las muchas obligaciones de responsabilidad proactiva.

Además de un tratamiento de datos, el uso público de IA podrá suponer también una decisión automatizada (art. 22 RGPD) con sus especiales garantías. Y como se dirá, en muchos casos el régimen y garantías de las discriminaciones especialmente prohibidas es fácil que converja y se superponga con la regulación del tratamiento de datos especialmente protegidos (art. 9 RGPD).

Igualmente y por último, no hay que descartar que las actuaciones del sistema de IA público discriminatorias, se enmarquen en contextos de infracción sancionadora o penal en su caso, particularmente sensibles y ajenos por lo general a la necesidad de que el hecho se enmarque en el procedimiento Administrativo.

### 3.2.- El grado de legalidad exigible para diseño, implantación y uso del sistema de IA público

Hay que contextualizar normativamente la existencia, usos y efectos del sistema de IA público y su relación con el principio de legalidad y con el ejercicio de potestades administrativas<sup>39</sup>. Como es sabido, el principio de legalidad es muy variable en sus intensidades y así habrá de serlo respecto del uso público de IA y todas las acciones tendentes a ello. Habrá de exigirse una mayor regulación de la posibilidad de uso de estos sistemas según su impacto y cuando pueda implicar actuaciones restrictivas de derechos, penales, sancionadoras o determinar la posibilidad de ejercer derechos. La exigencia de legalidad se debilita en los casos de usos de IA en actuaciones favorables o ampliatorias de derechos como en el contexto de prestación de servicios públicos, sanidad, educación, etc. En la misma dirección, habrá que intentar delimitar si el sistema de IA se utiliza para el ejercicio de potestades administrativas, pues la decisión de empleo, diseño y desarrollo del sistema de IA habrán de contar con variable cobertura legal y normativa. No obstante, no siempre la ley y las normas determinan expresamente la potestad administrativa en sí. La potestad puede ser implícita o inherente al cumplimiento de una competencia o atribución. Lamentablemente, es muy posible que las acciones más importantes para el diseño, configuración, implantación y evaluación del sistema de IA queden al margen del reconocimiento normativo y se consideren potestades implícitas o inherentes. Sin embargo, en la mayoría de los casos estas decisiones son los elementos determinantes para generar el sesgo o discriminación algorítmica. Así las cosas, el uso de la IA pública puede contar con una muy variable cobertura y delimitación normativa. Aquí se apuesta firmemente por el uso de los sistemas IA para el ejercicio de potestades discrecionales, precisamente como garantía frente a los humanos. Y, como luego se insiste, dado que estos usos de IA suponen más riesgos, debe exigirse una mayor habilitación normativa que incluya garantías. En esta línea hay que tener en cuenta el XVIII. 6 d) de la Carta de De-

Se proyecta para este ámbito GAMERO CASADO, E. y FERNÁNDEZ RA-MOS, S. (2016): Manual básico de Derecho Administrativo, (13ª ed.), Tecnos, 2016, págs. 77-78.

rechos Digitales que impone normativa con garantías adecuadas para las decisiones discrecionales automatizadas.

La exigencia de densidad legal normativa serán mayores según mayor sea el peligro e impactos en derechos e intereses, el mayor grado de automatización y participación humana, el peligro por la tecnología empleada. A lo anterior hay que añadir específicamente las variables propias al ámbito de la igualdad. Mayor cobertura normativa será necesaria especialmente si el sistema de IA establece diferencias a colectivos y especialmente a los sospechosos de discriminación. Aunque sin excesiva claridad, en el artículo 2.2° Ley 15/2022 exige una ley o disposición general). Asimismo, si el sistema de IA público trata datos personales, cosa que no debe ser nada extraña, la base de legitimación legal es la exigencia general (Informe 175/2018 AEPD) y sólo excepcionalmente el consentimiento puede servir como base de legitimación. Y como se expondrá infra esta exigencia de legalidad se hará muy intensa en los supuestos de que el sistema de IA trate datos especialmente protegidos, que es precisamente lo más habitual en los casos de supuestos especialmente sospechosos de discriminación.

#### 4.- ¿CUÁNDO EL USO PÚBLICO DE UN SISTEMA DE IA TIENE IMPACTO Y SUFICIENTE RELEVANCIA JURÍDICA Y PUEDE CONSTITUIR UNA DISCRIMINACIÓN ALGORÍTMICA?

## 4.1.- Los criterios y variables para determinar el impacto, nivel de riesgo y relevancia jurídica de los sistemas algorítmicos públicos

Cabe partir de un concepto inclusivo de sistemas públicos de IA que no se ciña estrictamente al concepto de "IA", sino que gire sobre el uso público de algorítmicos, esto es, sistemas informáticos que integren fórmulas más o menos complejas y las apliquen a los datos. Todos los sistemas algorítmicos potencialmente pueden generar sesgos y discriminaciones. Con acierto la Ley 1/2022, de 13 de abril valenciana es inclusiva cuando impone la publicidad activa "de sistemas algorítmicos o de inteligencia artificial", no sólo de estos últimos. Ahora

bien, los riesgos serán mayores conforme el sistema sea más complejo y especialmente debido a la autonomía del sistema propiamente de IA. Y en particular los riesgos se dan con los sistemas predictivos así como los sistemas de autoaprendizaje y aprendizaje profundo, por su falta de explicabilidad, que hará muy difícil justificar los tratos diferentes o demostrar que no hay sesgos relevantes.

Por cuanto a la atracción para el ámbito del Derecho, cabe seguir algunos criterios básicos, algunos de ellos ya están establecidos normativamente y lo que propongo es proyectarlos también para los sistemas IA público y en nuestro caso en relación con los sesgos y la discriminación.

Así, hay que tener en cuenta que el sistema de IA público "produzca efectos jurídicos en él o le afecte significativamente de modo similar" al individuo (art. 22 RGPD). Y acierta también la Ley 1/2022, de 13 de abril valenciana cuando un elemento esencial (para que deba o no hacerse publicidad activa de un sistema) es "que tengan impacto en los procedimientos administrativos o la prestación de los servicios públicos" (art. 16.1 l). Este concepto de "impacto", tan habitual en las evaluaciones de riesgos y en el contexto de la protección de datos, pasa a ser clave.

Para valorar el impacto hay que tener especialmente en cuenta, de un lado, la finalidad del uso o propósito del sistema de IA. Es un buen punto de partida seguir los usos de IA públicos de "alto riesgo" del futuro RIA, a partir de su artículo 6 combinado con los anexos II y III. En esta dirección y para el sector público considero que fundamentalmente serán sistemas de IA de alto riesgo los de seguridad para infraestructuras críticas públicas, el uso de dispositivos de salud, la gestión y priorización de emergencias, el uso de sistemas de identificación biométrica en el contexto de la seguridad pública y fronteras, los usos relacionados con la aplicación policial y judicial de la ley: evaluación de riesgos individuales de delinquir y reincidir, evaluación de perfiles, pruebas y polígrafos, detección estado emocional, análisis de documentación. Asimismo, en el Anexo III del RIA se consideran de alto riesgo usos de IA pública para determinar la admisión y programas en el ámbito de la educación, evaluación y seguimiento de aprendizaje. También los sistemas relativos a los servicios de empleo y evaluación de rendimiento y, más próximo a la actividad administrativa, los sistemas de evaluación, concesión, revocación de prestaciones y servicios esenciales de asistencia pública. Se dice que un tercio de los sistemas de IA públicos serán de alto riesgo<sup>40</sup>. Aunque son de alto riesgo de sesgos y discriminación todos los que están, ni mucho menos están todos los que son. Dejando ahora al margen los usos vinculados a defensa y seguridad nacional, me permito escalar de mayor a menor impacto y riesgo de discriminación, así como de atracción jurídica diversas categorizaciones de usos de IA públicos, a saber:

- Sistemas que identifican o priorizan objetivos para la aplicación de la ley o realizan inspecciones en el ámbito de infracciones penales, administrativas y persecución de ilicitudes y fraude, cada vez más habituales en ámbitos de seguridad, mercado y competencia, trabajo, salud, cuidado del medio ambiente, etc.
- Sistemas para la adjudicación de contratos, subvenciones, privilegios del gobierno.
- Sistemas para personalizar, priorizar o apoyar la prestación de servicios a los ciudadanos en salud, educación, empleo, servicios sociales, etc.
- Sistemas de extracción de información, investigación, recopilación, la supervisión y el análisis de datos, para la elaboración de políticas, la toma de decisiones, monitoreo general y análisis de riesgos.
- Sistemas para la gestión de la organización interna, recursos humanos y las adquisiciones y la gestión de los recursos tecnológicos.

JOINT RESEARCH CENTRE, TANGI, L. y otros (2022): AI Watch European landscape on the use of Artificial Intelligence by the Public Sector, JRC Science For Policy Report, Publications Office of the European Union, Luxembourg, 2022, p. 58 doi:10.2760/39336, https://publications.jrc.ec.europa.eu/repository/handle/JRC129301

También se afirma que en general, sólo un 10% de sistemas de IA serán de alto riesgo, COMISIÓN EUROPEA, RENDA A. (2021): Study to Support an Impact Assessment of Regulatory Requirements for Artificial Intelligence in Europe. Final Report (D5), abril 2021. https://op.europa.eu/es/publication-detail/-/publication/55538b70-a638-11eb-9585-01aa75ed71a1

 Sistemas para interactuar y comunicarse con el público sobre sus derechos y obligaciones y su participación<sup>41</sup>.

Señala Huergo que los sistemas de selección de sujetos para ser inspeccionados "quedarían en un segundo plano y con una "irrelevancia jurídica casi total" Por el contrario aquí se consideran del mayor impacto. Y es que pese a que las normas y jurisprudencia actuales en España parecen darle la razón, es momento de evitar esta huida del Derecho y aplicar las necesarias garantías a estos sistemas automatizados que encienden las luces rojas determinando o guiando la actuación humana. Los falsos positivos y negativos pueden generar tremendas discriminaciones. El caso judicializado más importante hasta la fecha de Países Bajos de SyRI, con mucho acierto determinó la nulidad del sistema que se utilizaba sólo en la fase de selección de alertas.

Para determinar el mayor o menor riesgo del sistema de IA público y la posibilidad de que haya generado o no el sesgo o la discriminación son importantes los *niveles de mayor o menor automatización e intervención o supervisión humana*. Siguiendo los estándares 2016 para el

recho administrativo", en HUERGO LORA, A. J. (dir.), DÍAZ GONZÁLEZ, G. M. (coord.) La regulación de los algoritmos, Aranzadi Thomson Reuters, Cízur,

2020, p. 77.

<sup>41</sup> Elaboración propia, combinando diversas descripciones de usos de IA públicos, en JOINT RESEARCH CENTRE, TANGI, L. y otros (2022), ob. cit. antes en JOINT RESEARCH CENTRE, MISURACA, G. (2020), AI Watch. Artificial Intelligence in public services. Overview of the use and impact of AI in public services in the EU, Joint Research Centre, Unión Europea. https://publications. irc.ec.europa.eu/repository/bitstream/JRC120399/jrc120399 misuraca-aiwatch\_public-services\_30062020\_def.pdf De especial interés, en EEUU destaca FREEMAN ENGSTROM, D. y otros (2020): "Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies. Report submitted to the Administrative conference of the United States", NYU School of Law, Public Law Research Paper No. 20-54, 122 págs., en concreto, pp. 21-69, http://dx.doi. org/10.2139/ssrn.3551505; HERMSTRÜWER, Y. (2020): "Artificial Intelligence and Administrative Decisions Under Uncertainty", en WISCHMEYER, T y RA-DEMACHER. T. (ed.), (2020), ob. cit. pp. 199-223. También, VELASCO RICO, C. I. (2022): "Tecnologías disruptivas en la Administración Pública. Inteligencia artificial y Blockchain", en CERRILLO I MARTÍNEZ, A. (coord.) La Administración Digital, Dykinson, Madrid, pp. 228-256, pp. 236-240. 42 HUERGO LORA, A. J. (2020): "Una aproximación a los algoritmos desde el de-

vehículo autónomo<sup>43</sup>, recientemente Roehl<sup>44</sup> ha afirmado seis niveles de automatización de la actuación administrativa: "automatización mínima"; "recuperación y tratamiento de datos"; cuando el sistema sugiere los "pasos procedimentales a seguir"; "decisiones asistidas"; "decisiones automatizadas", hasta el grado "decisiones autónomas" con sistemas dinámicos de machine learning no supervisado.

En este punto cabe recordar que las especiales garantías del artículo 22 RGPD sólo se dan respecto de las decisiones "únicamente" automatizadas del sector público. Igualmente, las escasas garantías del artículo 41. 1º LRJSP se reservan a "cualquier acto o actuación realizada íntegramente a través de medios electrónicos" y que "no hava intervenido de forma directa un empleado público". Hoy por hoy estos requisitos sirven para una huida del Derecho, esto es, en tanto en cuanto se considera que hay intervención humana en la decisión administrativa que se adopte, no se aplican las especiales garantías. No obstante, el Grupo del artículo 29 ha puesto límites, consciente del seguimiento rutinario por los humanos del sistema automatizado. Para que no rijan estas especiales garantías, la intervención humana ha de ser "significativa, en vez de ser únicamente un gesto simbólico" y llevada a cabo por "persona autorizada y competente". El estudio de impacto que debe hacerse ha de registrar el grado de intervención humana<sup>45</sup>. La Carta de Derechos Digitales exige el estudio de impacto también en las decisiones "semiautomatizadas" públicas (apartado XVIII. 4°). El futuro RIA para considerar si el sistema es de alto riesgo ha introducido en el artículo 6.3º

https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic

ROEHL, U. (2022): "Understanding Automated Decision-Making in the Public Sector: A Classification of Automated, Administrative Decision-Making", en Juell-Skielse, G. y otros (eds.), Service Automation in the Public Sector. Progress in IS. Springer, Cham, pp. 35-63, DOI: 10.1007/978-3-030-92644-1\_3, sigo por MORAL SORIANO, L. (2022): "Inteligencia artificial y filosofía del derecho decisiones automatizadas, derecho administrativo y argumentación jurídica", en LLANO ALONSO F. H. (dir.), Inteligencia artificial y filosofía del derecho, Laborun, Madrid, 2022, pp. 475-498, https://idus.us.es/bitstream/handle/11441/137250/Inteligencia%20artificial\_Llano%20Alonso.pdf?sequence=1&isAllowed=y

GRUPO DEL ARTÍCULO 29 (2018): Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679. 3 de octubre de 2017 y revisadas el 6 de febrero de 2018, p. 23.

de la versión de la República Checa que "la salida del sistema no es puramente accesoria con respecto a la acción o decisión pertinente que debe tomarse". Así pues, hay que valorar si la salida del sistema de IA público, por ejemplo, una propuesta de resolución, puede pasar a ser elemento sustancial para la decisión final que adopte el usuario de su sistema.

Acierta Huergo cuando señala que los mayores riesgos se dan respecto de los sistemas predictivos por cuanto inciden en la acción administrativa añadiendo un "contenido de elaboración propia"<sup>46</sup>. Por el contrario, otros sistemas tienen "un valor meramente auxiliar en la aplicación de la norma"<sup>47</sup> y cabe ubicarlos en el ámbito de la actuación reglada. Valero en 2007 y Martín Delgado en 2009 hablaban de los sistemas automatizados de "baja intensidad" frente a los de "discrecionalidad política o discrecionalidad técnica no parametrizable"<sup>48</sup>, que afirmaban que no eran posibles en la legislación española.

Sobre estas bases, considero que el peligro del uso de sistemas de IA públicos de mayor a menor pasaría por el uso meramente instrumental (procesadores, traductores, etc.); también meramente sería instrumental el uso para actos reglados, siempre que el sistema se limite a seguir la decisión normativa. Más riesgos se dan en el uso de IA pública en ámbitos de discrecionalidad técnica y más aún en el uso para ámbitos de discrecionalidad política o discrecionalidad técnica no parametrizable. También podrían añadirse a esta escala las decisiones IA administrativas para la aplicación de conceptos jurídicos indeterminados.

Ahora bien, pese a la utilidad de estas escalas de riesgo, hay que insistir que desde el nivel más básico de sistemas automatizados y algorítmicos es más que posible que el sistema genere sesgos, errores

<sup>&</sup>lt;sup>46</sup> HUERGO LORA, A. J. (2020), *ob. cit.* ver pp. 65 y ss. en particular 66 y 75 y ss.

<sup>&</sup>lt;sup>47</sup> *Ibídem*, p. 65)

VALERO TORRIJOS, J. (2007): El régimen jurídico de la e-Administración. El uso de los medios informáticos y telemáticos en el procedimiento administrativo común, 2.ª ed., Comares, Granada, pp. 74 y 75. Y MARTÍN DELGADO, I. (2009): "Naturaleza, concepto y régimen jurídico de la actuación administrativa automatizada", Revista de administración pública, nº 180, pp. 353-386. Ver pp. 368 y ss.

y discriminaciones. Basta una mala selección o calidad de datos o un mero error en un simple algoritmo o fórmula para que quede viciado e impacte en la decisión final que se adopte, pese a que sea formalmente humana.

# 4.2.- ¿Cuándo el sesgo y el error algorítmico es relevante para constituir una discriminación? Datos estadísticos en discriminaciones indirectas y la necesidad de "recalibrarlos" a la baja para la IA pública

Debe valorarse en cada supuesto específico si el sesgo o trato diferente que genera el sistema de IA tiene una relevancia o efectos suficientes para considerar que puede constituir una discriminación. Los errores, sesgos o posibles discriminaciones deben ser evaluados y determinarse hasta qué punto son asumibles por la organización<sup>49</sup>. Como recuerda Zlotnik hay que utilizar métricas de impacto y no sólo métricas de error. El impacto no es lo mismo que el error, dado que un 0,1% de errores de que una persona muera, se prorrogue su estancia en la cárcel, sea detenida o que se produzca un incidente en una central nuclear, nada tiene que ver con la posibilidad de que se le conceda erróneamente una ayuda escolar, se le priorice para una cita médica general o se le concedan entradas gratuitas para a una actividad cultural.

No es fácil establecer apriorísticamente los umbrales asumibles o no, pero sí es posible exigir que se hayan analizado, determinado, evaluado y adoptado medidas al respecto. Si estamos en el ámbito de diferencias de trato de las especialmente sospechosas o concurren otros derechos fundamentales en juego, serán menos tolerables los sesgos y errores.

Tiene interés señalar ahora que el juicio de si se da efectivamente una discriminación indirecta -que luego se expone-, se centra precisamente en la determinación de los "efectos sustancialmente más perjudiciales para un grupo protegido" <sup>50</sup>. En nuestro caso, se trata de valorar datos

<sup>&</sup>lt;sup>49</sup> ZLOTNIK, A. (2019), ob. cit. pp. 27-28.

<sup>&</sup>lt;sup>50</sup> STC 145/1991, FJ 2°; STC 128/1987, FJ 6°, ver voto particular a la STC 240/1999.

estadísticos, de la realidad, de las circunstancias concretas del caso en el que el sistema de IA -aparentemente neutro- afecta de manera negativa especialmente a determinados colectivos. El TJUE ha indicado que la diferencia de trato debe ser considerable y que el resultado perjudicial debe tener una incidencia de en torno al 80 o 90% para las personas pertenecientes al grupo protegido. Lo habitual no es centrarse en magnitudes numéricas, sino que "para presumirse discriminatoria, la medida debe afectar a un número *mucho más elevado* de mujeres que de hombres [Rinner-Kühn], o a un porcentaje *considerablemente inferior* de hombres que de mujeres [Nimz, Kowalska] o a un porcentaje *muy superior* de mujeres que de hombres [De Weerd]"<sup>51</sup>.

Estos parámetros serían en principio aplicables a los sistemas de IA. Sin embargo, considero que para sistemas de IA públicos que se aplican de forma masiva hay que "recalibrar" estos umbrales y ser mucho menos tolerantes. 52 Para establecer el umbral de afectación del sesgo o diferencia algorítmica en un caso concreto, no sólo hay que tener en cuenta la afectación al concreto afectado por un sistema de IA, sino que habrá de ponderarse el peligro que supone que ese error o sesgo masivo se replique en miles o millones de decisiones. Además, ha de tenerse en cuenta que si el error no se controla, analiza y en su caso se corrige, las decisiones erróneas pasarán a ser big data que alimentará a los futuros algoritmos, haciendo que el sesgo se cronifique. En consecuencia, en vez de considerar como relevantes diferencias de trato "muy elevadas", "considerablemente inferiores" o "muy superiores", habrán de bastar tasas de afectación mucho menores. Se seguiría así para el ámbito algorítmico la doctrina Sentencia TJUE 9 de febrero de 1997.53

Sobre el papel de las estadísticas y otros datos FRA- CONSEJO DE EUROPA (2018), *ob. cit.* pp. 275-282. Ahí las referencias a sentencias TIUE mencionadas.

Se aprovecha la idea del recalibrado de garantías de CITRON, D. K. (2008), *ob.cit.*, pp. 1254 y ss. p. 1286.

C-167/97, Regina contra Secretary of State for Employment, ex parte Nicole Seymour-Smith y Laura Perez, párrafo 61. Sigo por SORIANO ARNANZ, A. (2021 a).

#### 5.- DISCRIMINACIONES ALGORÍTMICAS POR ERROR, DIRECTAS, INDIRECTAS, INTERESECCIONALES Y ACCIONES POSITIVAS ALGORÍTMICAS

No procede extenderse ahora en la protección internacional, constitucional y legal de la igualdad y la no discriminación.<sup>54</sup> En España, la igualdad se ha conformado jurídicamente tanto como un valor superior del ordenamiento jurídico (art. 1. 1° CE), así como un principio y derecho fundamental; también su constitucionalización ha incluido las dos vertientes de la igualdad formal (art. 14 CE) y la igualdad material (art. 9. 2° CE), entre otros preceptos<sup>55</sup>. Baste simplemente adelantar que las leves o reglamentos en su contenido, así como quienes los aplican, tienen en general prohibido establecer diferencias de trato que no tengan una justificación objetiva y razonable. Y esta prohibición se hace más intensa si las diferencias se basan en circunstancias especialmente sospechosas o prohibidas de discriminación. En estos casos especiales, las diferencias de las normas o de su aplicación se someten a un escrutinio de justificación objetiva y proporcionalidad más riguroso y, de hecho, se presume que sí que son discriminatorias. Puede adelantarse también que en el ámbito del sector público

<sup>54</sup> Entre otros, destaco FRA- CONSEJO DE EUROPA (2018): Manual de legislación europea contra la discriminación, revisión 2018 340 pp., Oficina de Publicaciones de la Unión Europea, https://www.echr.coe.int/documents/handbook\_non\_discri\_law\_spa.pdf; GONZÁLEZ-TREVIJANO, P. (2020): Los principios de igualdad y no discriminación, una perspectiva de Derecho Comparado. España, EPRS Servicio de Estudios del Parlamento Europeo, PE 659.297, octubre, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/659297/ EPRS\_STU(2020)659297\_ES.pdf; RUBIO LLORENTE, F. (1993): "Igualdad en la jurisprudencia del Tribunal Constitucional. Introducción", en La Forma del Poder (Estudios sobre la Constitución), CEC, Madrid, pp. 637-684; GIMÉNEZ GLUCK, D. (2004): Juicio de igualdad y Tribunal Constitucional, J.M. Bosch Editor, Barcelona; REY MARTÍNEZ, F. (2017): Igualdad y prohibición de discriminación: de 1978 a 2018, Revista de derecho político, Nº 100, 2017, pp. 125-171 y REY MARTÍNEZ, F. (2020): Derecho antidiscriminatorio, Aranzadi Thomson Reuters, Cízur.

No agotan estos preceptos la recepción expresa de la igualdad en la Constitución, así, el artículo 23. 2º CE supone una concreta manifestación de la igualdad ante la ley y de la igualdad política. De otra parte, los artículos 139 y 149. 1.º. 1ª CE vienen a reconocer el carácter estatal de la igualdad en el marco de la descentralización autonómica.

el control de la igualdad queda muy ligado y se superpone con el del cumplimiento de otros mandatos constitucionales: la actividad administrativa no puede ser arbitraria, ha de estar motivada con suficiencia y racionalidad jurídica (arts. 9. 3 y 24 CE), ha de someterse a la legalidad y al mandato de objetividad (art. 103 CE) y el control incluirá que no se dé un desvío de poder (art. 106 CE). Como se verá, esta superposición se da más en los supuestos generales de igualdad en la aplicación de la ley , no en los especialmente sospechosos de discriminación.

A continuación se proyectan los variados conceptos y categorías básicas de la igualdad respecto del uso público de algoritmos. Y para ello es conveniente apoyarse en las recientes ente definiciones del artículo 6 Ley 15/2022.

## 5.1.- La discriminación o sesgo algoritmo será en la mayoría de supuestos "discriminación por error"

Aunque no sea un concepto asentado en el Derecho antidiscriminatorio, resulta especialmente útil acudir a la noción de "discriminación por error", esto es, "aquella que se funda en una apreciación incorrecta acerca de las características de la persona o personas discriminadas" (art. 6. 2° b Ley 15/2022). Este concepto permite subsumir prácticamente todos los supuestos en los que el sistema de IA público genera diferencias de trato por error, sin que se haya diseñado para establecer estas diferencias. Cabe adelantar que en estos casos no habrá de aplicarse el test de igualdad, sino que el análisis jurídico se centrará en probar la existencia del sesgo o discriminación, determinar si es relevante y qué consecuencias tiene y si se han cumplido las medidas preventivas para evitarlo.

## 5.2.- Discriminaciones algorítmicas directas, interseccionales y por asociación

Las discriminaciones directas (art. 6. 1° Ley Orgánica 3/2007, de 22 de marzo y art. 6. 1° Ley 15/2022)<sup>56</sup> se dan cuando una persona

<sup>&</sup>lt;sup>56</sup> FRA- CONSEJO DE EUROPA (2018), *ob. cit.* pp. 47-58.

recibe un trato menos favorable por alguna de las circunstancias especialmente sospechosas de discriminación. En el caso de un sistema de IA tendría lugar si se introducen en el sistema de IA datos sobre la pertenencia a un colectivo especialmente discriminado y a dicha pertenencia se asocia un factor negativo, o si los algoritmos y variables se diseñan para perjudicar a estos colectivos.

Ya se trate de unas u otras discriminaciones, los sistemas algorítmicos públicos también pueden generar discriminación múltiple o interseccional (art. 6. 3°) en tanto se acumulen errores o tratamientos diferenciados basados en dos o más circunstancias especialmente prohibidas (por ejemplo, mujeres musulmanas). Ello no será muy extraño dada la ingente cantidad de datos y perfiles que se manejan.

La categoría de la "discriminación por asociación" (art. 6. 2° a) Ley 15/2022) puede ser útil para abordar las diferenciaciones algorítmicas a partir de datos de aproximación, características sociales o *proxies*<sup>57</sup>. Como señala Soriano esta categoría "abre la puerta a la posibilidad de reconocer como discriminación algorítmica (directa o indirecta)" estos supuestos en los que se tienen en cuenta "afinidades" étnicas, multiculturales.

#### 5.3.- La complejidad de las discriminaciones algorítmicas indirectas

Es bastante habitual acudir a la categoría de las discriminaciones indirectas en el caso de algoritmos, aunque la cuestión en modo alguno resulta clara. Cabe recordar que se da una discriminación indirecta "cuando una disposición, criterio o práctica aparentemente neutros ocasiona o puede ocasionar a una o varias personas una desventaja particular con respecto a otras por razón de las causas" de las especialmente sospechosas. (art. 6. 2º Ley orgánica 3/2007, de 22 de

Cabe recordar que se define "discriminación por asociación cuando una persona o grupo en que se integra, debido a su relación con otra sobre la que concurra alguna de las causas" especialmente sospechosas, art. 6. 2 a) Ley 15/2022. Al respecto de este tipo de discriminación FRA- CONSEJO DE EUROPA (2018), ob. cit. pp. 56-58.

<sup>&</sup>lt;sup>58</sup> SORIANO ARNANZ, A. (2021 a), *ob. cit*.

marzo y art. 6. 1° b) Ley 15/2022)<sup>59</sup>. En nuestro caso, se acude a esta categoría por cuanto lo "aparentemente neutral" sería el sistema de IA en su conjunto, dado que está diseñado sobre la base de datos, criterios o variables como mínimo aparentemente objetivas<sup>60</sup>. Lo normal es que el sistema de IA no se haya diseñado para diferenciar, pero se hayan generado involuntariamente sesgos que dan lugar a una discriminación por error. Es por ello que el análisis jurídico en las discriminaciones indirectas acaba siendo bastante similar al caso de la discriminación por error, centrado esencialmente en determinar los efectos del sistema y el umbral de admisión de los sesgos, errores y diferencias que genera. No tiene sentido aplicar el test general de discriminación, puesto que no se trata de justificar si la diferenciación que produce el sistema está justificada y es razonable y proporcional<sup>61</sup>.

Aunque cae a medio caballo entre las discriminaciones directas e indirectas, puede ser útil acudir a las discriminaciones indirectas respecto de problemas del uso de *proxies*, como se verá *infra*, datos o variables que son "aparentemente neutros", pero de los que el sistema de IA acaba derivando un tratamiento perjudicial para colectivos especialmente discriminados. En estos supuestos también se puede acudir al concepto de discriminación por asociación algorítmica y para Soriano estos supuestos como abordables como discriminación directa<sup>62</sup>. A mi juicio, lo importante es apreciar que estos datos *proxie* que se manejan son afines a datos de colectivos especialmente discriminados, o bien, que es el sistema de IA al tratar con *proxies* el que produce los efectos diferentes para los colectivos especialmente

<sup>&</sup>lt;sup>59</sup> FRA- CONSEJO DE EUROPA (2018), *ob. cit.* pp. 58-77.

Afirma SORIANO ARNANZ, A. (2021 a), *ob. cit.* que una discriminación algorítmica indirecta se da por "la disposición, práctica o criterio aparentemente neutro puede ser la variable específica introducida en el sistema que, por el valor que le otorga este, genera un resultado discriminatorio para el grupo desaventajado, o bien el algoritmo generalmente considerado". Por cuanto a la variable específica será necesario acceder al sistema y la complejidad del mismo y de los datos que procesa dificultará especialmente el conocimiento.

SORIANO *ibídem*, acude erróneamente a mi juicio a este test en las discriminaciones algorítmicas indirectas, si bien, acaba reconduciendo su análisis a la necesidad de analizar la precisión del sistema.

<sup>62</sup> SORIANO ARNANZ, A. (2021 b) y (2022 a).

protegidos. Si es así, se han de aplicar en lo posible las garantías de las discriminaciones especialmente prohibidas que luego se siguen.

También, en principio, se estaría propiamente en el ámbito de la discriminación algorítmica indirecta cuando el sistema de IA público no diferencia en sus efectos y resultados a los colectivos especialmente discriminados, es decir, cuando arroja resultados iguales para categorías diferentes<sup>63</sup>. Este tratamiento algorítmico aparentemente neutral puede ser inaceptable. Este supuesto sería similar a los que se abordan en el siguiente apartado y deben ser analizados bajo el test de la igualdad y no tanto en el marco de las discriminaciones indirectas.

### 5.4.- Acciones positivas algorítmicas, deber de diferenciación y modulaciones y compensaciones para la minimización de la discriminación

El mandato constitucional de hacer efectiva la igualdad (art. 9.2° CE) ha de proyectarse también en la acción y la transformación digital del sector público. Entre otros, queda recogido en la Carta de Derechos Digitales, especialmente en su apartado VIII, así como los IX y XII respecto las luchas contra las brechas digitales, que incluye también "la promoción de "políticas públicas específicas dirigidas a abordar las brechas de acceso atendiendo a posibles sesgos discriminatorios" (2°)<sup>64</sup>. También como proyección del artículo 9.2° CE, para asegurar la igualdad de todas las personas en el disfrute de los derechos, las administraciones, los empleadores y los prestadores de servicios deben adoptar medidas especiales o específicas para adaptar sus disposiciones y prácticas a las personas con características diferentes, denominadas generalmente como acciones positivas<sup>65</sup>.

En sentido contrario a lo señalado por SORIANO *ibídem*.

Al respecto, ABA CATOIRA, A. (2022), "Derechos de igualdad, personas con discapacidad y mayores en el entorno digital (VIII, XI Y XII)", COTINO HUESO, Lorenzo (editor), La Carta de Derechos Digitales, Tirant Lo Blanch, Valencia, pp. 123-154.

Sobre el tema, entre otros, FRA- CONSEJO DE EUROPA (2018), ob. cit. pp. 78-91. La bibliografía es muy abundante, además de obras generales mencionadas, BALAGUER CALLEJÓN, M. L. (2018): "Acciones positivas para la igualdad efectiva de mujeres y hombres", VENTURA FRANCH, A. y GARCÍA

Asimismo, aunque es un tema conflictivo, el TEDH afirma desde 2000 que se vulnera de derecho de igualdad "cuando [...] los Estados no aplican un trato diferente a personas cuyas situaciones son sensiblemente diferentes" puede ser obligatoria una regulación o trato diferente en algún caso concreto para salvar el derecho de igualdad<sup>67</sup>.

En consecuencia, el sistema de IA público en ocasiones habrá de diseñarse para instrumentar una medida especial o acción positiva, esto es, con la intencionada finalidad de beneficiar a determinados grupos sociales que están en situación desfavorable respecto de la media de la población. Se trata de medidas algorítmicas "orientadas a prevenir, eliminar y, en su caso, compensar cualquier forma de discriminación o desventaja en su dimensión colectiva o social" (art. 6.7° Ley 15/2022, también, art. 11 Ley Orgánica 3/2007, de 22 de marzo), al amparo del artículo 9.2 CE o el artículo 23 de la Carta de derechos fundamentales de la UE. Si el sistema y su funcionamiento está regulado, habrá que analizar la concreta disposición y la justificación objetiva razonable de la medida de acción positiva y su concreta proporcionalidad. No obstante y en especial, habrá que tener en cuenta las medidas concretas efectivamente adoptadas en el sistema de IA. En estos supuestos se modulará el funcionamiento del sistema y se introducirán los criterios y variables para preferenciar a los colectivos determinados, o perjudicar al resto, según se mire. Habrán de reco-

CAMPÁ. S. (coord.) El derecho a la igualdad efectiva de mujeres y hombres: una evaluación del primer decenio de la Ley Orgánica 3/2007, pp. 285-312; GIMÉNEZ GLUCK, D. Una manifestación polémica del principio de igualdad: acciones positivas moderadas y medidas de discriminación inversa, Tirant lo Blanch, Valencia, 1999; GONZÁLEZ-TREVIJANO, P. (2020), ob. cit. pp. 75-77.

Especialmente importante fue la sentencia TEDH, Thlimmenos contra Grecia [GS], n.o 34369/97, 6 de abril de 2000, apartado 44. De manera análoga, TEDH, Pretty contra Reino Unido, n.o 2346/02, 29 de abril de 2002, apartado 88.

TEDH, Biao contra Dinamarca [GS], n.o 38590/10, 24 de mayo de 2016, apartado 103; TEDH, D. H. y otros contra República Checa [GS], n.o 57325/00, 13 de noviembre de 2007, apartado 184. Ciertamente, nuestro TC siempre ha afirmado lo contrario, esto es, que no hay un derecho a la diferenciación (por todas, sentencia 181/2000 (FJ 14°) o STC 128/2004, de 21 de julio). Sobre el tema, de escasa atención, COBREROS MENDAZONA, E. (2997): "Discriminación por indiferenciación: estudio y propuesta", *Revista española de derecho constitucional*, Año n° 27, N° 81, 2007, págs. 71-114.

gerse las evidencias relativas a los datos de entrenamiento del sistema y las variables utilizadas, así como los mecanismos de evaluación del funcionamiento del sistema con los criterios correctores introducidos. Sobre estas evidencias, será posible analizar las medidas respecto del sistema de IA público y valorar si se han adoptado conforme a la normativa -si es que la hay- y ya en particular, la justificación objetiva<sup>68</sup>. Ciertamente se trata de evaluar la legitimidad de los objetivos seguidos y la proporcionalidad de los medios utilizados por el sistema de IA para alcanzarlos.

Cabe señalar que es un análisis similar al que hay que realizar respecto de las correcciones, medidas de compensación, modulación o mitigación que se han de adoptar precisamente en el sistema de IA cuando se detecta la discriminación y el sesgo algorítmicos y hasta dónde hacerlo.

#### 6.- ¿CUÁNDO UN SISTEMA ALGORÍTMICO PÚBLICO QUE NO TIENE EN CUENTA CIRCUNSTANCIAS ESPECIALMENTE SOSPECHOSAS ES DISCRIMINATORIO?

# 6.1.- La igualdad "en" la ley o norma que habilite el uso del sistema de IA y el control laxo de la posibilidad de establecer diferencias de tratos razonables y no arbitrarios

La igualdad formal pasó a incluir también la igualdad *en* la ley, esto es, la necesidad de que las normas respeten también en su contenido el principio de igualdad<sup>69</sup>. La acción normativa legal o reglamentaria implica otorgar -o no- el mismo tratamiento a grupos o colectivos definidos por la norma. Bajo esta perspectiva, la

<sup>68</sup> Sobre la justificación FRA- CONSEJO DE EUROPA (2018), *ob. cit.* p. 103-125.

La STC 144/1988, de 12 de junio. En su FJ 1º se afirma que "frente al legislador o frente al poder reglamentario, impidiendo que uno u otro puedan configurar los supuestos de hecho de la norma de modo tal que se dé trato distinto a personas". Diversas referencias para el TC de la igualdad en la ley, GONZÁLEZTREVIJANO, P. (2020), ob. cit. p. 55-58.

función de las normas es atribuir efectos jurídicos diferentes a las situaciones jurídicas que crean. El diseño, implantación y uso del sistema de IA público queda sometido al principio de legalidad de la Administración, de modo más o menos intenso según diversas variables que se han expuesto supra. Sobre esta base, es fundamental tener en cuenta la cobertura legal o reglamentaria del sistema de IA y de la posibilidad de que dicha normativa habilite a que el sistema de IA genere diferencias de trato. Así, por ejemplo, la norma puede determinar requisitos para una contratación, una prestación, una subvención y, en consecuencia, el diseño, configuración y datos del sistema de IA quedarán más o menos recogidos en la norma. Si es así, habrá que llevar a cabo un análisis de igualdad en dicha norma. En este sentido, la igualdad opera "frente al legislador o frente al poder reglamentario, impidiendo que uno u otro puedan configurar los supuestos de hecho de la norma de modo tal que se dé trato distinto a personas que, desde todos los puntos de vista legítimamente adoptables, se encuentran en la misma situación". Así pues, no hay que dar "relevancia jurídica a circunstancias que, o bien no pueden ser jamás tomadas en consideración por prohibirlo así expresamente la propia Constitución, o bien no guardan relación alguna con el sentido de la regulación que, al incluirlas, incurre en arbitrariedad y es por eso discriminatoria" (STC 144/1988, de 12 de junio FJ 1°)<sup>70</sup>. La igualdad -y también la igualdad en la ley- no sólo se predica respecto de las discriminaciones especialmente sospechosas. Ahora bien, lo cierto es que salvo que se trate de discriminaciones prohibidas o vinculadas con otro derecho fundamental, el análisis de admisibilidad se hace muy laxo en general, limitado a la razonabilidad, no arbitrariedad y existencia de justificación objetiva. En muchos casos basta con que el establecimiento de la diferencia tenga una finalidad y ésta sea razonable, básicamente que no sea contraria al sistema de valores

Y, "para que las diferenciaciones normativas puedan considerarse no discriminatorias resulta indispensable que exista una justificación objetiva y razonable, de acuerdo con criterios y juicios de valor generalmente aceptados, cuya exigencia deba aplicarse en relación con la finalidad y efectos de la medida considerada, debiendo estar presente, por ello, una razonable relación de proporcionalidad entre los medios empleados y la finalidad perseguida" STC 46/1999, de 22 de marzo FJ 2°. También clara STC 75/1983, de 18 de febrero".

constitucionales. En principio es necesario que el trato sea congruente y proporcionado<sup>71</sup>, si bien desaparece -o casi- el análisis de la proporcionalidad. Los términos del análisis de la posible discriminación quedan abiertos al decisionismo judicial y a los criterios políticos<sup>72</sup>. También es cierto que en el ámbito de la potestad reglamentaria y frente a la arbitrariedad se van incrementando las necesidades de justificar la finalidad y contenido de la norma reglamentaria, esencialmente en el procedimiento de elaboración de la misma<sup>73</sup>. Por el contrario, el análisis sí será riguroso respecto de las normas que habiliten a que un sistema de IA pueda establecer diferencias de trato por motivos especialmente sospechosos de discriminación.

Así, STS de 3 de noviembre de 2010, recurso 480/2009 y la más reciente STS de 4 de abril de 2022 (rec.300/2020), https://delajusticia.com/wp-content/uploads/2022/04/STSregla.pdf

<sup>&</sup>quot;La diferencia de trato a personas en situaciones sustancialmente similares [...] es discriminatoria si no tiene una justificación objetiva y razonable; en otras palabras, si no persigue un fin legítimo o no hay una relación de proporcionalidad razonable entre los medios empleados y el fin perseguido" (TEDH, Burden contra Reino Unido, 29 de abril de 2008, ap. 60). FRA- CONSEJO DE EUROPA (2018), ob. cit. pp. 103-108 y ss. Así desde la STEDH de 23 de julio de 1968 (caso régimen lingüístico belga), asumida por el TC en el FJ 3º de la pronta STC 22/1981, de 2 de julio.

GIMÉNEZ GLUCK, D. (1999), *ob. cit.* p. 34 ha llegado a afirmar en esta dirección que "En materia de constitucionalidad de la diferencia en el trato, las decisiones son principalmente políticas, aunque se revistan de un entramado jurídico justificativo de la opción política que se adopta. No es menos cierto que esto es lo que ocurre con la mayoría de las decisiones denominadas jurídicas, pero en el terreno de la igualdad esta dinámica general se hace mucho más evidente." (Cfr. *Una manifestación polémica del principio de igualdad...* cit. pág. 34). "La discrecionalidad de las decisiones del Tribunal, que en general ya puede ser considerada excesiva, alcanza en la determinación de la constitucionalidad del trato desigual un grado ilimitado. No obstante, la construcción jurisprudencial sobre la justificación objetiva y razonable otorga, al menos, un marco jurídico a estas decisiones políticas" (pp. 38 -39).

# 6.2.- Control de discriminación y arbitrariedad y exigencias de motivación en los casos de aplicación algorítmica de la ley. Decisiones automatizadas discrecionales y supuestos no especialmente sospechosos

La ley o reglamento es muy posible que no establezcan ninguna diferencia de trato, pero que sea el sistema de IA el que la produce. En estos supuestos, ya no se trata de una cuestión de igualdad *en* la ley, sino de igualdad *ante* la ley o en la aplicación de la misma. La Administración y los jueces deben aplicar la ley por igual a quienes estén en la misma situación. La igualdad no exige un estricto trato igual, sino que exige que el diferente trato no sea discriminatorio. Pues bien, ello mismo debe predicarse de los sistemas IA públicos que tengan efectos jurídicos sobre las personas. Si el sistema de IA y la diferencia de trato cuenta con cobertura normativa, ya habrá tenido que realizarse en abstracto el análisis de la posible discriminación *en* la ley.

A partir de ahí, habrá que apreciar si las diferencias de trato que genera el sistema son relevantes en los términos ya vistos. Si es así, procede distinguir si el trato diferenciado se genera por sesgos y errores del sistema. Si es así, por lo que se trataría de una discriminación por error. En estos casos, el debate jurídico se centrará en una cuestión de prueba de la existencia del sesgo o discriminación, de umbrales de precisión admisibles, del cumplimiento de las medidas preventivas en el diseño, implantación y evaluación del sistema y de las consecuencias del sesgo o discriminación. No habrá de aplicarse el juicio de igualdad: si el sesgo, error o trato diferenciado es relevante y no cuenta con cobertura normativa o el sistema se ha desviado de la norma, no podrá considerarse justificado, razonable y no arbitrario.

Otra situación es que las normas no hayan previsto la diferencia de trato, pero en el diseño e implantación del sistema se han adoptado conscientemente decisiones para que el sistema de IA genere tratos diferentes. También puede darse el caso de que el sistema autónomo, especialmente de aprendizaje automático, funciona generando diferencias de trato no previstas. En estos casos, habrá que hacer un análisis del concreto trato diferente que depara el sistema, su posible finalidad objetiva y los efectos reales del sistema de IA. Así, no sólo habría que estar a las finalidades perseguidas o declaradas por los desarrolladores, sino a los resultados efectivos del sistema. Y a partir de ahí, para

que se considere que el sistema no es discriminatorio es preciso que la finalidad del sistema y sus resultados puedan considerarse razonables y no contrarios al sistema de valores, a la Constitución, así como al marco legal de la actuación administrativa, así como congruentes y proporcionados. Como se adelantó, el juicio de igualdad es más laxo cuando no se trata de una discriminación especialmente prohibida (por ejemplo, que el sistema preferencia quienes estudiaron en colegios privados). Ahora bien, en estos supuestos *el análisis ha de concurrir llegando a fusionarse con el de no arbitrariedad, de motivación con suficiencia y racionalidad jurídica* (arts. 9. 3 y 24 CE) exigibles a la actuación administrativa.

En los supuestos en los que los sistemas automatizados más simples se emplean en decisiones regladas, el análisis será más sencillo y simplemente se limitaría a apreciar si el sistema y los algoritmos se ajustan completamente a la norma que lo regula.

Las curvas se dan especialmente si se usa un sistema de IA para decisiones y ejercicio de potestades discrecionales<sup>74</sup>. Es en estos ámbitos donde el uso de IA público más puede aportar para lograr mayor neutralidad y objetividad, reducir la discriminación y la discrecionalidad y los ineludibles sesgos y opacidad humanos, además de lograr mayor justicia con la personalización de individualización de los casos<sup>75</sup>. De hecho, a mi juicio, no tiene sentido una prohibición general del uso de IA respecto de potestades discrecionales, como se interpreta habitualmente la actual normativa alemana. Eso sí, el uso en estos contextos discrecionales tiene un peligro potencial mayor y, en consecuencia, mayores deben ser las garantías que deben darse. Es así como debe interpretarse el derecho de la Carta (XVIII. 6° d): "Que la adopción de decisiones discrecionales quede reservada a personas, salvo que normativamente se prevea la adopción de decisiones automatizadas

Se sigue a continuación una proyección de los conceptos y garantías exigibles al respecto en GAMERO, E. y FERNÁNDEZ, S. (2016), *ob. cit.* pp. 81-84.
 En esta línea, MINISTERIE VAN BINNENLANDSE ZAKEN (2021), *ob. cit.*

p. 11. TISCHBIREK, A. (2020), *ob. cit.* pp. 103 y ss.; HERMSTRÜWER (2020), *ob. cit.* pp. 199 y ss. En cierto modo, HUERGO LORA, A. J. (2021), *ob. cit.* Y especialmente habría que acudir a la excelente reflexión global de BEN- SHAHAR, O. y PORAT, A. (2021): *Personalized law: different rules for different people*, Oxford University Press.

con garantías adecuadas." Y precisamente la labor normativa debe ser más intensa en estos casos. En las decisiones discrecionales la Administración y el sistema de IA pueden optar entre diversas soluciones hacia la más favorable al interés general. En estos casos se ha de poder comprobar que el sistema de IA no es arbitrario y el trato diferente tiene justificación objetiva y no es irrazonable. El sistema de IA habrá de garantizar que sus decisiones están claramente respaldadas por los hechos determinantes sobre los que incide la decisión administrativa. Y para ello la importancia de la motivación administrativa es esencial, y para un sistema de IA este examen pasa por la suficiente transparencia y explicabilidad del sistema, especialmente relacionada con la motivación<sup>76</sup>. Como he analizado<sup>77</sup>, acierta la Carta de Derechos digitales al reconocer el derecho a "Obtener una motivación comprensible en lenguaje natural de las decisiones que se adopten en el entorno digital, con justificación de las normas jurídicas relevantes, tecnología empleada, así como de los criterios de aplicación de las mismas al caso" (XVIII. 6° c). Hay que eludir, por ello, una presunción jurídica de acierto de las decisiones automatizadas o de sistemas inteligentes y ofrecer posibilidades y garantías eficaces —no sólo en la vía judicial— de probar los posibles errores de diseño o funciona-

COTINO HUESO, L. (2022): "Derechos ante la administración digital y la inteligencia artificial (XVIII Y XV)", en COTINO HUESO, Lorenzo (editor) (2022) ob. cit. pp. 251-284.

<sup>76</sup> Al respecto, una mayor atención en HERMSTRÜWER, Y. (2020), ob. cit.; GA-MERO CASADO, E. (2021): "Necesidad de motivación e invalidez de los actos administrativos sustentados en inteligencia artificial o en algoritmos", Almacén de Derecho, 4 febrero, https://almacendederecho.org/necesidad-de-motivacione-invalidez-de-los-actos-administrativos-sustentados-en-inteligencia-artificial-oen-algoritmos Desde la perspectiva de la discrecionalidad, HUERGO LORA, A. J. (2021): "Administraciones Públicas e inteligencia artificial: ¿más o menos discrecionalidad?", El Cronista del Estado Social y Democrático de Derecho, nº 96-97, 2021, pp. 78-95, sigo por https://laadministracionaldia.inap.es/noticia. asp?id=1512494; CERRILLO MARTÍNEZ, A. (2020): "¿Son fiables las decisiones de las Administraciones públicas adoptadas por algoritmos?", European Review of Digital Administration & Law Volume 1 - Issue 1-2 - 2020, DOI: 10.4399/97888255389603, págs. 18-36, ver 23-26, y 29 y ss.. Con relación a la revisión judicial ulterior, en especial, DE LA SIERRA MORÓN, S. (2020): "Inteligencia artificial y justicia administrativa: una aproximación desde la teoría del control de la Administración Pública", Revista General de Derecho Administrativo, nº. 53, 2020.

miento. En el caso de la discrecionalidad técnica, será necesario que se conozcan las fuentes de información sobre las que el sistema de IA realiza el juicio técnico, los criterios de valoración cualitativa que el sistema ha empleado, y que el resultado de salida del sistema permita motivar la aplicación de tales criterios al resultado individualizado que ha preferenciado una decisión o solución. El sistema de IA debe poder hacerlo, así como generar evidencias de cómo lo hace al tiempo de motivar su decisión.

Estas obligaciones de motivación están muy ligadas a la justificación objetiva atenuada, que es lo que esencialmente se exige en los supuestos de discriminaciones no especialmente prohibidas<sup>78</sup>. En principio, habría de haber cierta tolerancia de errores en estos supuestos de uso de IA en ámbitos de discrecionalidad técnica.

Propuse en la fase inicial de la Carta<sup>79</sup> -y afortunadamente se ha incluido- la garantía de que "El interesado tendrá derecho a que se motive o se explique la decisión administrativa cuando esta se separe del criterio propuesto por un sistema automatizado o inteligente" (XVIII. 6° c). El interesado debe poder conocer las razones humanas que imponen a no seguir las propuestas o decisiones automatizadas, se trata de una extensión necesaria precisamente del control o supervisión humana.

Las exigencias al sistema de IA serán bastante similares cuando se utilicen para la interpretación de conceptos jurídicos indeterminados en supuestos concretos por la Administración. En estos casos, para actuar conforme a Derecho se da un control de racionalidad y razonabilidad. Es por ello que el sistema de IA que se aplique debe permitir que se controle si la Administración ha respetado la realidad de los hechos, si ha considerado todos los factores relevantes y si en sus criterios y preferencias ha seguido el valor que el ordenamiento confiere a estos factores. Respecto de la transparencia y explicabilidad me remito al trabajo en esta obra, así como a recientes aportaciones.<sup>80</sup>

<sup>&</sup>lt;sup>78</sup> FRA- CONSEJO DE EUROPA (2018), ob. cit. pp. 103-108 y ss.

Véase mi aportación en esta misma dirección en RED DAIA (2020 a): Aportación a consulta de derechos digitales de julio, 2020 https://acortar.link/q1wl3i

<sup>(2023): &</sup>quot;Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida", Revista Española de Transparencia y "Transparencia y explicabilidad de la inteligencia artificial y "compañía" (comunicación, interpre-

### 7.- ¿CUÁNDO UN SISTEMA ALGORÍTMICO PÚBLICO QUE SÍ TIENE EN CUENTA CIRCUNSTANCIAS ESPECIALMENTE SOSPECHOSAS ES DISCRIMINATORIO?

# 7.1.- Las garantías frente a las discriminaciones algorítmicas especialmente prohibidas se superponen por lo general a las garantías de los tratamientos de datos especialmente protegidos

El derecho a la igualdad excluye toda clase de discriminación, si bien hay una serie de rasgos diferenciadores que por la experiencia de la Historia y el presente son especialmente sospechosos de motivar una discriminación. El artículo 14 CE así como su interpretación conforme a los tratados y jurisprudencia internacionales parecen a garantizar la igualdad en general y a incluir un mandato antidiscriminatorio específico, una prohibición específica de discriminar "por razón de nacimiento, origen racial o étnico, sexo, religión, convicción u opinión, edad, discapacidad, orientación o identidad sexual, expresión de género, enfermedad o condición de salud, estado serológico y/o predisposición genética a sufrir patologías y trastornos, lengua, situación socioeconómica, o cualquier otra condición o circunstancia personal o social". De igual modo, cabe tener en cuenta las diferencias de trato en razón de la "nacionalidad, de si son menores o mayores de edad o de si disfrutan o no de residencia legal" (art. 2.1° Ley 15/2022), así como el régimen más específico de los tratos diferentes en razón de enfermedad (art. 2. 3°). En la normativa europea, las características protegidas se limitan expresamente al sexo, el origen racial o étnico, la edad, la discapacidad, la religión o las creencias, y

tabilidad, inteligibilidad, auditabilidad, testabilidad, comprobabilidad, simulabilidad...). Para qué, para quién y cuánta", en COTINO HUESO L. Y CASTE-LLANOS J. (coords.), *Transparencia y explicabilidad de la inteligencia artificial*, Tirant lo Blanch, Valencia, 2022. Una clara visión de conjunto de interés en JIMÉNEZ LÓPEZ, Jesús (2022): "Oscuridad algorítmica en el sector público", en VESTRI, G. (2022): *Disrupción tecnológica en la administración pública retos*, Aranzadi, Cízur.

la orientación sexual<sup>81</sup>. Pues bien, cuando se dan estas circunstancias, se incrementan y mucho las garantías frente a la no discriminación. Y lo que es más, en muchos casos a estas garantías hay que añadir las -muchas- de protección de datos en general y de los datos especialmente protegidos en particular. La superposición de los dos regímenes lleva a una solución bastante más garantista que la que sólo otorga el artículo 14 CE, especialmente para el sector público.

### 7.2.- En la mayoría de casos será necesaria una ley de calidad y con garantías que habilite a que un sistema de IA pueda diferenciar en razón de datos especialmente sospechosos de discriminación. Y no la hay

En primer lugar, hay que analizar si la norma que habilita a la actuación administrativa y al uso del sistema de IA recoge expresamente la posibilidad de tener en cuenta circunstancias especialmente sospechosas y en qué medida (datos de recogida, variables a tener en cuenta por el sistema, etc.). Como se vio, las exigencias del principio de legalidad administrativa en principio son más intensas debido al impacto en derechos. Por ello, el punto de partida es que la norma sí que debería recoger expresamente esta posibilidad. De hecho, el artículo 2.2° Ley 15/2022 parece exigir "norma con rango de ley, o [...] disposiciones normativas o decisiones generales de las administraciones públicas." Pero es más, en razón de la aplicación de la normativa de protección de datos en muchos supuestos sí que es obligatoria una norma con rango de ley -y no reglamentaria-. Además, no se tratará de la general base legal que se exige para que sector público trate datos personales (y no el consentimiento, Informe 175/2018 AEPD). Y es que el tratamiento de datos relativos a categorías especialmente sospechosas de discriminación coincide mucho con el tratamiento de datos especialmente protegidos del artículo 9 RGPD82. Además, en diversos casos confluirá

FRA- CONSEJO DE EUROPA (2018), *ob. cit.* p. 182, ver pp. 177-260.

<sup>&</sup>quot;origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o datos relativos a la vida sexual o las orientación sexuales".

con el artículo 22 RGPD para decisiones automatizadas, que también impone ley83. A la vista del antedicho art. 2.1° Ley 15/2022, únicamente no estaríamos en supuestos de tratamientos de datos especialmente protegidos cuando el sistema de IA trate datos de "nacimiento", género, "edad", "lengua" o "situación socioeconómica" y en algunos de estos casos, como la edad o el género no sería difícil que se exijan especiales garantías de protección de datos. Cuando sí estemos bajo el régimen especial del artículo 9 RGPD no basta cualquier ley, sino que la misma ha de contar con una especial calidad según STC 76/2019, de 22 de mayo, en especial FJ 8° y, por supuesto, con las exigencias del artículo 23 RGPD. En el contenido de estas leves claramente se tiene que habilitar a la Administración y legitimar el tratamiento concreto del sistema de IA y además justificar el mismo. Esta justificación habrá de conectarse por lo general, en aras de un "interés público esencial" (art. 9. 2º g) RGPD). Asimismo, la jurisprudencia constitucional ha sido muy rigurosa respecto de la necesidad de que la ley integre en su contenido no sólo el detalle de la restricción y sus presupuestos, sino que también se han de regular las garantías concretas compensatorias para estos tratamientos de datos especialmente protegidos (y sospechosos de discriminación).

Resulta ciertamente difícil pensar que algún sistema público de IA en España actualmente cumpla con estos requisitos. Lo más cercano sería la Ley 22/2018, de 6 de noviembre, de la Generalitat de la Comunidad Valenciana que reguló bastante concienzudamente un sistema de alertas tempranas que en principio no manejaba datos sensibles y lo cierto es que el desconocido informe de la Informe AEPD 385661/2017 de marzo sobre el mismo no consideró suficiente la regulación<sup>84</sup>. Como excepción, es bastante positivo el artículo 10.5° (y Considerando 44) RIA, que habilita al uso de datos especialmente

La conexión de estos regímenes se aborda en cierto modo Consejo de Europa (2018), *op.cit.*, p. 38, así como SORIANO ARNANZ, A. (2021 a).

Al respecto dediqué mi estudio COTINO HUESO, L. (2021). Ver AEPD (2017), Informe AEPD 385661/2017 de marzo relativo al anteproyecto de ley de sistema de alertas tempranas de la Comunidad Valenciana. No publicado. He dispuesto un enlace en https://acortar.link/AJ6mYz

Asimismo, el 29 de julio de 2020 ejercí un derecho de acceso a la información pública con 17 preguntas sobre el sistema, resuelto favorablemente el 20

protegidos, precisamente, para evitar los sesgos y la discriminación. Algunas normas en España tienen en cuenta la necesidad de base de legitimación para la gestión de datos en sistemas de personalización del sector público, fenómeno ampliamente estudiado por Cerrillo. Es positivo el artículo 19, Ley 18/2020, de 28 de diciembre, de facilitación de la actividad económica sobre proactividad de la Administración, de Cataluña. Si bien, ni siquiera sería suficiente para datos especialmente protegidos. En el ámbito de la proactividad, las meras referencias a la personalización de servicios en modo alguno valdrían como base de legitimación legal<sup>85</sup>. Los artículos 5 a), 25. 1° y 67. 3° Ley 4/2019, de 17 de julio de Administración digital de Galicia. También el artículo 28.6° Ley 8/2019, de 5 de abril, para una Administración más ágil, de Extremadura, es bastante general. Con ser positiva, tampoco sería suficiente al menos para datos especialmente protegidos la referencia del artículo 50.3° Ley 5/2021, de 29 de junio, de organización y régimen jurídico del sector público autonómico de Aragón. El Decreto 76/2020, de 4 de agosto, de Administración digital de Cataluña, (que obviamente no es ley) tiene acertadas referencias a la proactividad y la personalización (at. 4), que por sí no serían suficientes como habilitación, ni tampoco la mención a las garantías (art. 31). Cuestión diferente es que parece tratarse de usos automatizados de bajo impacto, una personalización limitada a la "finalidad de informar a las personas, de manera predictiva y anticipada, sobre los servicios públicos a los cuales pueden acceder" (art. 31.1°). Y por lo general el consentimiento al que acude el artículo 25.3° Ley 4/2019, de 17 de julio de Administración digital de Galicia solo valdría para legitimar servicios personalizados puramente informativos y relativamente instrascendentes o que el ciudadano tenga la alternativa a los mismos.

de agosto (GVRTE/2020/1163199). He dispuesto un acceso a la resolución que incluye las preguntas formuladas en https://bit.ly/37f2apR

El tema lo aborda convenientemente aunque no como aquí desde el punto de vista de los datos especialmente protegidos o las discriminaciones especialmente prohibidas. CERRILLO I MARTÍNEZ, A. (2022): "La personalización de servicios digitales", en CERRILLO I MARTÍNEZ, A. (coord.) *La Administración Digital*, Dykinson, Madrid, pp. 311-370, pp. 333 y ss.

#### 7.3.- Un canon de control "mucho más" estricto, e incluso excepcionalmente estricto

La norma legal o en su caso reglamentaria que habilite a que el sistema de IA genere tratos diferenciados especialmente prohibidos, habrá de ser sometida a un juicio de igualdad estricto. El artículo 2.2º Ley 15/2022 permite tratos diferentes en estos supuestos específicos si son razonables y objetivos, con propósitos legítimos, también finalidades de proteger a personas o grupos de población. Asimismo, en razón del artículo 4.2º "No se considera discriminación" el trato diferente "que pueda justificarse objetivamente por una finalidad legítima y como medio adecuado, necesario y proporcionado para alcanzarla." Hay que advertir que la superación de estos estándares habrá de mirarse con lupa: "el canon de control, al enjuiciar la legitimidad de la diferencia y las exigencias de proporcionalidad resulta mucho más estricto, así como más rigurosa la carga de acreditar el carácter justificado de la diferenciación" (entre otras muchas, sentencia 75/1983, de 3 de agosto, FFJJ 6 y 7). Y ello se agravará en la mayoría de los casos en los que se traten datos especialmente protegidos en los que aún rige con mayor rigor la excepcionalidad y la prohibición. Así pues, en principio serán pocos los supuestos en los que no será necesaria una ley de calidad y muchas garantías para que un sistema de IA público trate datos especialmente protegidos y lo haga para establecer diferencias de trato.

En consecuencia, estos casos en los que hay ley pasarán a ser en principio una cuestión de igualdad en la aplicación de esa ley. Y en todo caso habrán de controlarse y bajo estos estándares estrictos todas las medidas de gestión de datos, diseño del sistema y técnicas y organizativas que se hayan adoptado para que el sistema de IA sí que establezca diferencias de trato.

En EEUU la clave jurídica para admitir que los algoritmos y el big data incluyan información racial reside en que el algoritmo no haya elegido el factor raza con finalidad discriminatoria ("because of" "discriminatory purpose"), que es lo que prohíbe la doctrina *Mc-Cleskey v. Kemp* 481 U.S. 279, 291-92 (1987). Por el contrario, quizá se podría de admitir que se tome en cuenta el factor raza por ser un elemento que aumenta la precisión "a pesar de" ("in spite of") sus efectos sobre un grupo identificable. Habrá que analizar rigurosamen-

te cada caso concreto el funcionamiento y resultados de la IA, eso sí, bajo la presunción de discriminación.

### 7.4.- La necesidad de regular legalmente el uso de datos especialmente protegidos para poder evaluar y minimizar la posible discriminación

Como se recuerda en Países Bajos<sup>86</sup>, precisamente para contrarrestar los efectos involuntarios e indeseables de los algoritmos, puede ser necesario tratar datos personales especialmente protegidos como variables de control para el desarrollo de modelos y métodos. De este modo se pueden eliminar sesgos y corregir las deficiencias ocultas de los conjuntos de datos o los modelos. En la misma dirección, señala Soriano que "es más sencillo enseñar a los sistemas automatizados a no discriminar con base en las categorías sospechosas si dichos datos se introducen y consideran expresamente". 87 Si en el sector privado el consentimiento en muchos casos podría permitir el uso de estos datos para esta finalidad, no es así como regla general para el uso de IA en el sector público, por lo que se hace necesaria una especial regulación habilitante y con garantías. En esta línea, el RIA en su Considerando 44 afirma que "los proveedores deben poder tratar también categorías especiales de datos personales, como cuestión de interés público sustancial [...] con el fin de garantizar el seguimiento, la detección y la corrección de los sesgos en relación con los sistemas de IA de alto riesgo" y el artículo 10. 5º RIA contiene la habilitación expresa.

De particular interés los criterios generales afirmados en el documento MINIS-TERIE VAN JUSTITIE EN VEILIGHEID (Ministerio de Justicia y Seguridad), Rijksoverheid (Gobierno central) (2021), Richtlijnen voor het toepassen van algoritmen door overheden en publieksvoorlichting over data-analyses (Pautas para la aplicación de algoritmos por parte de los gobiernos y educación pública sobre análisis de datos), Directiva (Richtlijn), de 08-03-2021, https://acortar. link/6C226N, pág, 23 y ss. https://acortar.link/6C226N Se remite especialmente a LAMMERANT, H. y otros (2019): Big data decision-making processes and shortcuts of discrimination, NTM/NJCM-bull. 2018/1, p. 10-11 y ejemplos como van der SANGEN, M. (2019), Research on fair algorithms for policy, 6-8-2019, https://www.cbs.nl/nl-nl/corporate/2019/32/research-on-fair-algorithmsfor-policy

En la misma línea SORIANO ARNANZ, A. (2021 a), ob. cit.

## 8.- ZONAS GRISES. GARANTÍAS ANTE EL USO POR EL SISTEMA DE IA PÚBLICO DE "PROXIES" Y GRUPOS ALGORÍTMICOS

### 8.1.- Tratamientos diferentes a partir de "proxies" o datos afines o indirectos de grupos especialmente protegidos

Hay que tener en cuenta el fenómeno de los datos llamados proxies o indirectos, esto es, datos en principio no vinculados a circunstancias especialmente sospechosas de discriminación, ni datos especialmente protegidos (por ejemplo, datos geográficos, de pautas y productos de consumo, formas de vida, etc.). Recuerda Watcher cómo los dueños de perros, los adolescentes tristes, los jugadores de videojuegos, los padres solteros o los pobres se utilizan habitualmente para asignar recursos y tomar decisiones como qué publicidad mostrar, precio ofrecer o servicio público financiar88. En principio, no aplicarían las garantías más intensas. Sin embargo, fácilmente el sistema de IA puede inferir de estos datos factores especialmente prohibidos o datos especialmente protegidos. Es por ello que hay que evitar que se eludan intencionalmente las garantías que corresponderían y analizar posibles "enmascaramientos" 89. Apunta Van der Sloot en esta línea que los datos no personales pueden convertirse en datos sensibles en una fracción de segundo, que basar la protección jurídica en la naturaleza de los datos "no es la mejor manera de avanzar". 90 Como he

WACHTER, S. (2022): "The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law", 15 de febrero, *Tulane Law Review*, pp. 5-6, http://dx.doi.org/10.2139/ssrn.4099100

Entre otros, TERRY NICOLAS, "Big Data Proxies and Health Privacy Exceptionalism", en *Health Matrix n°* 24, pp. 65-108 (2014). Sobre el enmascaramiento de datos especialmente prohibidos, BAROCAS, S. y SELBST, A. D. (2016), *ob. cit.* pp. 692-693. NICOLAS, T.(2014) "Big Data Proxies and Health Privacy Exceptionalism", en *Health Matrix n°* 24, pp. 65-108. Para el ámbito de la justicia, SIMMONS, R. (2016): "Quantifying Criminal Procedure: How to Unlock the Potential of Big Data in Our Criminal Justice System", *Mich. St. L. Rev.* 947 (Ohio State Public Law Working Paper No. 362. pp. 977 y ss. http://dx.doi. org/10.2139/ssrn.2816006

VAN DER SLOOT, B. (2021): "Regulating non-personal data in the age of Big Data", en TZANOU, M. (Ed.) (2021): Health data privacy under the GDPR:

sostenido al respecto en el ámbito de salud, el tratamiento jurídico ha de venir claramente ligado con la finalidad de su uso y el riesgo que se genera<sup>91</sup>. De este modo, hay que poner especialmente el foco en la afinidad de los datos con datos sensibles o de categorías sospechosas de discriminación. O, directamente, analizar si los efectos que genera el tratamiento impactan en estos colectivos sospechosos de discriminación, aunque los datos no sean especialmente sensibles. Si es así, no puede considerarse que el sistema sea aparentemente neutro y habrá que aplicar las especiales garantías de las discriminaciones prohibidas. Como he adelantado, en estos supuestos puede resultar práctico acudir a las categorías de discriminación por asociación y especialmente la discriminación indirecta. Desde EEUU, se afirma que algunos *proxies* como la información demográfica si se identifican "deben evitarse" por su "alto riesgo de conducir a la discriminación algorítmica" <sup>92</sup>.

### 8.2.- Creación dinámica de grupos y colectivos humanos y la privacidad y no discriminación colectiva y de grupo

Una de las novedades que implica la IA es la creación dinámica de grupos y colectivos humanos que no comparten en principio rasgos sospechosos. Apunta Watcher que la IA también crea grupos aparentemente incomprensibles, definidos por parámetros que desafían la comprensión humana, como píxeles en una imagen, comportamiento de clic, señales electrónicas o tráfico web y a resultas de ellos se deciden préstamos o solicitudes de empleo, que tienen un impacto significativo en la vida de las personas<sup>93</sup>. Ello además sin que los interesados percibamos ni la inclusión en estos grupos, ni las implicaciones que

Big Data challenges and regulatory responses, Routledge pp. 85-105; p. 16 de la versión SSRN.

OOTINO HUESO, L. (2020 b): "El alcance e interactuación del régimen jurídico de los datos personales y big data relacionado con salud y la investigación biomédica", Revista de derecho y genoma humano: genética, biotecnología y medicina avanzada, nº 52, pp. 57-96, nº52 enero-junio 2020. acceso

<sup>&</sup>lt;sup>92</sup> WHITE HOUSE (2022), *op.cit.*, p. 23 y 26.

<sup>&</sup>lt;sup>93</sup> WACHTER, S. (2022), *ob. cit.* pp. 6, 10, 24.

puede tener. Para Jaume,<sup>94</sup> en estos casos la identidad de los individuos no es relevante y se trata de la creación de grupos fuera de las categorías sociales convencionales de una cultura, e incluso al margen de los grupos tradicionalmente discriminados. Floridi y el grupo de Tilburg, que abordan especialmente la "privacidad de grupo en la era de los datos"<sup>95</sup>; exponen -y así lo recuerda también la AEPD<sup>96</sup>-, que la mayoría de las personas no están perfiladas individualmente, sino como miembros de un grupo específico. Los datos empleados para generar el perfil de estos colectivos o grupos podrían no tener la consideración de datos personales. Y afirma la AEPD que estos grupos "no están establecidos como tal y no están sujetos a ninguna forma jurídica", por ello concluye que el perfilado y las inferencias sobre estos grupos "carecerían de cualquier protección legal."<sup>97</sup>

Pues bien, como ya he sostenido<sup>98</sup> ciertamente no lo considero así en tanto en cuanto estas clasificaciones, aunque no supongan tratamiento de datos personales, pasan a serlo en el momento en el que se aplican a una persona concreta. En muchos casos serán supuestos de tratamiento de *datos inferidos*<sup>99</sup>, a los que Palma ha prestado especial

<sup>94</sup> JAUME PALASÍ, L. (2020): "Cómo la inteligencia artificial está impactando en las sociedades", Cerrillo I Martínez, A. Y Peguera Poch, M. (coords.), (2020): Retos jurídicos de la inteligencia artificial, Aranzadi, Cizur, pp. 27-39, p. 28.

En particular, KAMMOURIEH, L y otros, "Group privacy in the age of big data", en TAYLOR. L.; VAN DER SLOOT, B.; FLORIDI, L. (eds.) (2017): Group Privacy: New Challenges of Data Technologies, Springer, Capítulo 3, pp. 48-83, ver p. 54. https://www.stiftung-nv.de/sites/default/files/group-privacy-2017-authors-draft-manuscript.pdf

AEPD (2020 b): Privacidad de grupo, 19 de octubre, https://www.aepd.es/es/ prensa-y-comunicacion/blog/privacidad-de-grupo
 Ibídem.

<sup>(2022): &</sup>quot;Nuevo paradigma en la garantías de los derechos fundamentales y una nueva protección de datos frente al impacto social y colectivo de la inteligencia artificial", en COTINO HUESO, Lorenzo (editor), Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas, Thompson-Reuters Aranzadi, FIADI (Federación Iberoamericana de Asociaciones de Derecho e Informática), Cizur.

Especialmente cabe seguir, VAN DER SLOOT, B. (2021), ob. cit. y en España a POLO ROCA, A. (2021):, "Datos, datos, datos: el dato personal, el dato no personal, el dato personal compuesto, la anonimización, la pertenencia del dato y otras cuestiones sobre datos". Estudios de Deusto: revista de la Universidad de Deusto, Vol. 69, n°. 1, pp. 165-194, DOI: 10.18543/ed-69(1)-2021

atención<sup>100</sup>, esto es, datos de salida del sistema de IA a partir del tratamiento de datos personales aportados por el interesado o vinculados al interesado. Aunque el TJUE no es tan claro, sin duda alguna para el Grupo del Artículo 29 estos datos son datos personales, aunque no hayan sido facilitados el interesado<sup>101</sup>. En consecuencia, si el sistema de IA público confiere tratos diferentes en razón de la pertenencia a estos colectivos, como mínimo se aplicarán las garantías ordinarias de la discriminación, así como de la protección de datos. Y si puede considerarse que hay parámetros afines a los datos especialmente protegidos o a las categorías especialmente sospechosas, nos situaríamos ante un supuesto de *proxies* antes expuesto. Las soluciones propuestas por Watcher y su "teoría de la inmutabilidad artificial" y la proyección de su el "derecho a inferencias razonables" pese al interés de su reflexión, en cierto modo son reconducibles a la solución aquí expuesta.

### 9.-¿CÓMO PROBAR LA DISCRIMINACIÓN ALGORÍTMICA PÚBLICA? INVERSIÓN GENERAL DE LA CARGA DE LA PRUEBA Y LA NECESIDAD DE LA TRANSPARENCIA Y EXPLICABILIDAD

En general la discriminación se enfrenta a graves problemas de prueba y más en los casos de la discriminación algorítmica<sup>103</sup>.

Para empezar, lamentablemente hoy por hoy, ni siquiera se conoce la existencia misma de buena parte de los sistemas IA públicos, por lo

Sobre el tema, PALMA ORTIGOSA, A. (2021): Régimen jurídico de la toma de decisiones automatizadas y el uso de sistemas de inteligencia artificial en el marco del derecho a la protección de datos personales, Tesis doctoral Universidad de Valencia, 2021, que tuve el gusto de dirigir. Sobre la noción, WACHTER, S. y MITTELSTADT, B. D. (2019): "A right to reasonable inferences: re-thinking data protection law in the age of big data and AI", Columbia Business Law Review, vol. 2019, No. 2, pp. 494-620, p. 2 versión https://ssrn.com/abstract=3248829
 GRUPO DEL ARTÍCULO 29 (2018), ob. cit. p. 10.

WACHTER, S. y MITTELSTADT, B. D. (2019), ob. cit. entre otras pp. 49-50.

En general y por todos, respecto de la prueba de la discriminación, FRA- CON-SEJO DE EUROPA (2018), ob. cit. pp. 261-282. Respecto de algoritmos TISCH-BIREK, A. (2020), ob. cit.

que se hace más difícil saber si discriminan. Si se conocen, por lo general no se tratará de discriminaciones algorítmicas directas y evidentes en las que la norma o los responsables y proveedores del sistema de IA expresamente indiquen que en su diseño y medidas adoptadas han manejado el uso de datos de colectivos especialmente discriminados o que las variables del sistema preferencian o perjudican a unos colectivos frente a otros y en qué modo concreto. Si así es, obviamente será más sencillo probar la posible discriminación. De igual modo, especialmente por cuanto al funcionamiento autónomo del sistema de IA o de *machine* o *deep learning*, no será fácil que el mismo sistema muestre cómo efectúa tratos diferentes no previstos en el diseño. Y obviamente, si la causa de la discriminación son errores, éstos difícilmente serán plenamente visibles y parametrizados para que se puedan conocer y probar.

La tendencia del Derecho antidiscriminatorio en general ha sido la reducción de la necesidad o la inversión de la carga de la prueba por parte de la persona afectada. La regulación europea y española de no discriminación es bastante incisiva en terreno de la prueba. A continuación se señalan los elementos clave proyectados para el ámbito específico del uso de sistemas públicos de IA.

# 9.1.- Los indicios que debe aportar la persona afectada por el sistema de IA público. Inversión general de la carga de la prueba y sus excepciones

La persona afectada por un sistema de IA público (o la persona legitimada para su impugnación) debe aportar "indicios fundados" sobre la existencia de la discriminación (art. 30. 1º Ley 15/2022, de 12 de julio). Así pues, se requieren datos, documentación, estadísticas u otras pruebas sobre el sistema de IA que permitan hacer pensar que genera errores o resultados diferentes respecto de grupos o colectivos vinculados a categorías especialmente prohibidas, o diferentes niveles de falsos negativos o positivos. O que el sistema tiene en cuenta datos o variables relacionadas con estos colectivos. En algunos supuestos de discriminación indirecta, se tratará de mostrar indicios de que se trata de modo idéntico a quienes habrían de ser tratados de modo diferente por el sistema.

Señala Soriano que en principio no ha de ser problemático aportar términos de comparación del funcionamiento del sistema, al apreciar los resultados que arroja el sistema en otros muchos supuestos <sup>104</sup>. Asimismo, en muchos supuestos se podrá contar con información sobre el sistema.

En los casos de discriminaciones especialmente prohibidas, si se logra aportar suficientes indicios fundados, pasará a presumirse que el sistema de IA sí que es discriminatorio. Se invierte la carga de la prueba: "la carga de demostrar el carácter justificado de la diferenciación recae sobre quien asume la defensa de la misma" (sentencia 81/1982, de 21 de diciembre, FJ 2)<sup>105</sup>. Es al responsable del sistema de IA público a quien compete demostrar que no es discriminatorio. Ello sucedería en los supuestos de discriminaciones algorítmica directa, indirectas y encubiertas, así como discriminaciones por error. Esta inversión de la carga de la prueba también habrá de darse cuando el uso del sistema de IA quede sea vinculable al ejercicio de un derecho fundamental (salud, por integridad y vida, uso con, profesores, centros educativos, alumnos o sus padres educación, así como con sindicatos partidos, asociaciones ideológicas o religiosas, etc.).

En principio, en los supuestos en los que el sistema de IA público no parezca comprometer o afectar a colectivos tradicionalmente discriminados, ni quede vinculado con otros derechos fundamentales, no habría inversión de la carga de la prueba y el afectado habría de probar que el sistema contiene errores o sesgos relevantes o establece diferencias de trato que son claramente irrazonables. Sin embargo, considero que una vez probado el indicio de discriminación del sistema de IA público por la parte afectada, también habrá de darse en general la inversión de la carga de la prueba y que sea el responsable del sistema de IA público quien pruebe que no hay discriminación, sesgos y errores relevantes. A mi juicio debe reconocerse esta

Así SORIANO ARNANZ, A. (2022 a y b), se apoya en XENIDIS, R. y SEN-DEN, L. (2020): "EU Non-discrimination law in the era of artificial intelligence: mapping the challenges of algorithmic discrimination", en BERNITZ, U. y otros, (eds.), General Principles of EU Law and the EU Digital Order, Alphen aan den Rijn, Kluwer Law International, 2020, p. 171..

Al respecto, véanse las STC 81/1982, de 21 de diciembre, 34/1984, de 9 de marzo, 20/1985, de 14 de febrero, entre otras.

inversión general de la carga de la prueba en razón del llamado *data due process* desarrollado en buena medida en EEUU desde hace una década<sup>106</sup>. Y es que desde el punto de vista de los afectados por un sistema público de IA su posición es de una fuerte asimetría informativa frente al sector público y sus proveedores de los sistemas de IA. Ya sólo reunir "indicios" puede ser una labor extremadamente compleja.

Sin embargo, no será necesaria esta inversión de la prueba si no se da esta asimetría. No será necesaria si la diferencia de trato la establece la propia normativa que regula el sistema de IA o se trata de decisiones expresas en el diseño e implantación del sistema y sobre las mismas hay publicidad activa con suficiente información y explicabilidad del sistema de IA para la prueba de la discriminación.

Para que el afectado pueda hacerse con estos indicios o en su caso pruebas, pasa a ser esencial la existencia y cumplimiento de obligaciones de registros públicos de algoritmos (para algunos sistemas de alto riesgo del futuro RIA), de transparencia activa obligatoria de los sistemas IA utilizados (p. e. art. 16.1 l) Ley 1/2022, de 13 de abril), o las autorizaciones del uso de estos sistemas y su difusión en sede electrónica (p.e. arts. 11 y 13 RAFESP)<sup>107</sup>, la publicidad de los estudios de impacto que se efectúen, o del registro de tratamientos de datos (art. 31 Ley Orgánica 3/2018). Asimismo, puede ser clave el ejercicio del derecho de acceso a la información pública (arts. 20 CE y 105 CE, Ley 19/2013) o las de acceso de protección de datos y transparencia (arts. 15 y 12 y ss. RGPD), reforzado en el caso de decisiones automatizadas (art. 22 y 13 RGPD). Al respecto cabe remitir a este estudio, así como a diversos trabajos sobre las obligaciones de transparen-

Especialmente CRAWFORD, K. Y SCHULTZ, J. (2014): "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms". Boston College Law Review, Vol. 55, No. 93 https://ssrn.com/abstract=2325784.; CITRON, D. K. (2008), ob.cit. No expresan esta inversión general de la carga de la prueba, pero sí asentaron las obligaciones de generar registros, poder auditar los sistemas, obligaciones de información mínima, etc.

Sobre este reglamento y su regulación al respecto, puede seguirse mi estudio (2021 b): "El nuevo reglamento de Administración electrónica, que no innova en tiempos de transformación digital", *Revista Catalana de Derecho Público* nº 63,, pp. 118-136, DOI: http://dx.doi.org/10.2436/rcdp.i63.2021.3672

cia<sup>108</sup>. El seguimiento de políticas de transparencia no obligatorias por proveedores o usuarios de los sistemas de IA públicos puede ser también esencial para facilitar que los afectados conozcan la existencia del sistema, del impacto e importancia de este, su diseño, datos de entrada y funcionamiento.

Tampoco se daría esta inversión general de la carga en el uso de IA contextos en los que se da intensamente la presunción de inocencia: "procesos penales ni a los procedimientos administrativos sancionadores, ni a las medidas adoptadas y los procedimientos tramitados al amparo de las normas de organización, convivencia y disciplina de los centros docentes" (art. 13. 2° Ley Orgánica 3/2007, art. 30. 3° Ley 15/2022, de 12 de julio).

Hay que llamar la atención que en septiembre de 2022 se ha presentado una Propuesta de Directiva sobre responsabilidad en materia de IA<sup>109</sup>. La misma es especialmente destacable en materia de carga de la prueba y presunción de responsabilidad, en particular su artículo 4. Salvo en el ámbito penal, no hay que excluir su proyección respecto de la responsabilidad por uso público de la IA, así como en el caso de la colaboración público-privada.

### 9.2.- Lo que debe probar el proveedor o el usuario del sistema de IA público

En general, a partir de la aportación de los indicios suficientes por el afectado, es el responsable del sistema de IA público quien debe probar que no discrimina. Así, según los casos, el responsable del sistema habría de demostrar:

La literatura ya es abundante, me remito a los trabajos de la monografía CO-TINO, L. Y CASTELLANOS, J. (coords.) (2022), el ya referido de JIMÉNEZ LÓPEZ, Jesús (2022) ob.cit., en la misma obra, BENÍTEZ PALMA, E. "Mucho ruido y pocas nueces: algunas consideraciones sobre los registros públicos de algoritmos", pp. 83-96.

https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:52022PC0496 Sobre el tema en la UE

 $https://ec.europa.eu/info/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence\_en\\$ 

- Que el sistema no es erróneo ni genera sesgos o que estos son irrelevantes.

- Que la diferencia de trato se basa en un factor objetivo no relacionado con un motivo especialmente prohibido.
- Que el tratamiento diferente que genera el sistema cuenta con "justificación objetiva y razonable, suficientemente probada, de las medidas adoptadas y de su proporcionalidad" (art. 30. 1° Ley 15/2022, de 12 de julio)<sup>110</sup>. Así, el proveedor y en su caso la Administración como usuaria del sistema de IA habrían de expresar y detallar la finalidad o finalidades pretendidas y cómo se ha configurado el sistema, la selección de datos de entrada y entrenamiento y otros elementos, y que estas decisiones técnicas y organizativas arrojan resultados proporcionales respecto de tales finalidades y con relación a las diferencias que generan.
- Que, en su caso, si el sistema trata de modo más favorable a determinados colectivos y lo hace de forma razonable o proporcional, en supuestos de acciones positivas, obligaciones de diferenciación o medidas de minimización de sesgos, por ejemplo.

Puede ser muy importante para el responsable del sistema de IA -presunto discriminador- contar con las evidencias sobre el cumplimiento de todas las exigencias proactivas y preventivas, tanto las que se exijan debido a la protección de datos (como los estudios de impacto, análisis de riesgos, etc.) o de otra normativa aplicable, como en su caso, el futuro reglamento de IA. El cumplimiento de estos deberes genera toda una serie de información y documentación. Asimismo, la demostración del cumplimiento de estos deberes prueba la diligencia debida, así como la falta de cualquier ánimo discriminatorio o finalidad contraria a la ley. En su caso, las evidencias que puedan aportarse serán reactivas, esto es, a partir del conflicto y cuestionamiento del funcionamiento del sistema. La falta de esta información podrá ser la prueba de la discriminación por "inacción" (art. 4.1° Ley 15/2022) u omisión.

Igualmente cabe tener en cuenta, en los mismos términos que esta última ley, el artículo 77. 3º bis LPAC, de 1 de octubre, del Procedimiento y artículo 60. 7º Ley 29/1998, de 13 de julio, reguladora de la Jurisdicción Contencioso-Administrativa.

La regulación de igualdad también indica que el órgano judicial, a instancia de parte, podrá recabar, si lo estimase útil y pertinente, informe o dictamen de los organismos públicos competentes (art. 13. 1º Ley Orgánica 3/2007, art. 30. 2º Ley 15/2022, de 12 de julio). A lo anterior cabe aplicar todo lo procedente sobre las facultades administrativas o judiciales de solicitud de prueba, proyectadas al ámbito tecnológico y a la especialidad de los sistemas de IA en los que pueden concurrir importantes cuestiones de propiedad o secreto industrial, secreto y materia clasificadas u otras restricciones.

### 10.- MÁS VALE PREVENIR QUE DISCRIMINAR. ¿CUÁLES SON LOS REMEDIOS Y MEJORES PRÁCTICAS PARA EVITAR LA DISCRIMINACIÓN ALGORÍTMICA PÚBLICA?

Desde hace años instituciones y expertos vienen señalando los mejores remedios y buenas prácticas para evitar los sesgos, errores y discriminación. A continuación se destilan los que se consideran más relevantes. Hay que adelantar que buena parte de ellos no sólo son buenas prácticas o recomendaciones, sino que pasarían a integrarse en las medidas técnicas y organizativas que las autoridades de protección de datos ya exigen, que por ejemplo se han de adoptar en los estudios de impacto según los riesgos. Muchas medidas también se integrarían entre las muchas obligaciones preventivas que impone el futuro RIA respecto de un sistema de alto riesgo.

Desde 2018 el Grupo del Artículo 29 G29-UE<sup>111</sup> ha detallado una serie de medidas respecto decisiones automatizadas esencialmente "para garantizar que las personas reciben un trato justo y no discriminatorio", sin "resultados discriminatorios, erróneos o injustificados", entre otras insistía en los controles periódicos de calidad, auditorías algorítmicas, en su caso independientes de terceros, garantías contractuales respecto de los algoritmos de terceros que garanticen comprobaciones y cumplimiento normativo. Asimismo, mecanismos de certificación para operaciones de tratamiento, códigos de conducta

GRUPO DEL ARTÍCULO 29 (2018), ob. cit. pp. 37-38.

para procesos de auditoría, comités de ética para evaluar los daños y beneficios.

También desde 2018 y adoptadas finalmente en 2019, las Directrices para la ética en el diseño en la UE del HLEG<sup>112</sup> incluyen una lista de evaluación, un *checklist* con 150 cuestiones y existe también una adecuación para el ámbito de salud (www.future-ai.eu). Muchas de ellas son de interés para lograr la exactitud y fiabilidad (bloque 2)<sup>113</sup>, las integridad y calidad de los datos (bloque 3) y, específicamente, un bloque sobre "Diversidad, no discriminación y equidad" (5) con 19 cuestiones para evitar el sesgo y la discriminación<sup>114</sup>. También se incluyen cuestiones sobre accesibilidad y diseño universal y los bloques sobre bienestar social y ambiental y de responsabilidad también incluyen cuestiones de interés<sup>115</sup>. Se trata de una herramienta básica para la evaluación y autoevaluación de sistemas IA que perfectamente puede ser proyectada para el sector público, al igual que sucede con la Guía de auditoría de IA de la AEPD<sup>116</sup>.

HLEG (2019), *op.cit.*, pp. 26-31. De interés el portal ALTAI https://futurium.ec.europa.eu/en/european-ai-alliance/pages/welcome-altai-portal

<sup>¿</sup>Ha implementado medidas para garantizar que los datos utilizados sean completos y estén actualizados ?; ¿se necesitan datos adicionales?, ¿qué daño se causaría si el sistema de AI hace predicciones inexactas?, ¿el sistema está haciendo una cantidad inaceptable de predicciones inexactas ?; ¿pasos para aumentar la precisión del sistema ?, etc.

Entre otras, ¿Estableció una estrategia o un conjunto de procedimientos para evitar crear o reforzar un sesgo injusto en el sistema AI, tanto en el uso de los datos de entrada como en el diseño del algoritmo ?, diversidad y representatividad en los datos; inclusión poblaciones específicas, uso de pruebas en las fases de desarrollo, incorporación de mecanismos específicos para que otros puedan identificar problemas de sesgo o discriminación; mecanismos para consultar estos problemas; consideración de afectados indirectos. Evaluación si variación de resultados en mismas condiciones, cómo medirlo. ¿Cómo se ha medido la imparcialidad?

El bloque sobre bienestar social y ambiental (6) cuenta con alguna cuestión sobre si, por ejemplo, se tuvo en cuenta el impacto social general, como pérdidas de empleo. Finalmente, en el bloque de responsabilidad (7) también incluye cuestiones sobre auditoría independiente, evaluación de riesgo o impacto, capacitación y educación, introducción de mecanismos de "junta de revisión ética de AI" o similares.

AEPD (2021): Requisitos para Auditorías de Tratamientos que incluyan IA, 2021.

En cualquier caso, *supra* se han expuesto las causas de los sesgos y discriminaciones algorítmicas, agrupadas respecto de los datos, la prueba evaluación y validación de los sistemas y métricas y la consecuente adopción de medidas de mitigación y los factores humanos. Sobre estos ejes cabe recordar los principales remedios.

En el *primer grupo de remedios, relativos a los datos*, el NIST o la Casa Blanca<sup>117</sup> señalan que los datos definen en gran medida la tarea de aprendizaje automático en sí misma, por lo que hay que hay que "ir donde están los datos", una cultura centrada en qué conjunto de datos podría ser el más adecuado y no tanto en cuáles son accesibles. Emplear una representación estadística equilibrada en los conjuntos de datos con técnicas como medidas de desequilibrio de clase.

Tras la relevante sentencia de nulidad del sistema de IA SYRI, Países Bajos han destacado en Europa -con el Reino Unido- en la adopción de medidas preventivas y de transparencia algorítmica como ningún otro país<sup>118</sup>. Además de su estudio de impacto algorítmico IAMA<sup>119</sup>, destaca una guía específica para la no discriminación<sup>120</sup>. Se insiste en que hay que investigar la calidad de las fuentes de datos y si son suficientes para el propósito para el que se utilizan; identificar si existen limitaciones en el uso de una fuente de datos, algoritmo o método de análisis. Se recuerdan los requisitos legales de calidad, cantidad y actualización de

Destaca especialmente NIST; SCHWARTZ, R. y otros (2022), *op.cit.*, pp. 14-19 o WHITE HOUSE (2022), *op.cit.*, pp. 23-29, en especial pp. 26 y ss.

Destacan sus directrices MINISTERIE VAN JUSTITIE (2021), *ob. cit.*, para no discriminación y sesgos en especial pp. 23-24. Asimismo el Excel desarrollado por el ALGEMENE REKENKAMER (Tribunal de Cuentas), (2020): *Digitaal Toetsingskader Algoritmes (Marco de evaluación digital de Algoritmos)*, Tribunal de Cuentas, Excel, https://www.rekenkamer.nl/binaries/rekenkamer/documenten/publicaties/2021/01/28/download-het-toetsingskader/Toetsingskader+al goritmes+v1.0.xlsx

MINISTRY OF THE INTERIOR AND KINGDOM (2022): Impact Assessment. Fundamental rights and algorithms (IAMA), Países Bajos, marzo. https://www.government.nl/binaries/government/documenten/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms/Fundamental+Rights+and+Algorithms+Impact+Assessment.pdf

MINISTERIE VAN BINNENLANDSE ZAKEN; VAN DER SLOOT, B. y otros (2021): Non-discriminatie by design, 69 pp., encargo para (Ministerio del Interior), Tweede Kamer (Cámara de representantes), https://www.tweedekamer.nl/downloads/document?id=2021D22772

los datos, siendo los necesarios para los fines del análisis. Hay que describir la finalidad del sistema, cómo se manejarán los datos y los errores que se detecten<sup>121</sup>. Por cuanto a los datos de entrenamiento, hay que justificar la elección y poder parametrizar los posibles factores discriminatorios y análisis sobre las consecuencias de una elección diferente de la parametrización <sup>122</sup>.

Desde EEUU en 2022 se insiste en que los datos sean "representativos y sólidos". Y el uso de los *proxies*, como la información demográfica, "debe evitarse" por su "alto riesgo de conducir a la discriminación algorítmica". Deben realizarse pruebas proactivas para identificar los *proxies* y si se identifican eliminarse o buscar alternativas<sup>123</sup>. Respecto la prueba, evaluación, validación y verificación de los sistemas y los algoritmos y las métricas empleadas, el NIST aporta toda una guía frente a los sesgos, necesidades de modelación del sistema, correlaciones espurias, la complejidad algorítmica, validación del rendimiento, también el aseguramiento en el despliegue y el sesgo de implementación, la necesidad de garantizar actualizaciones periódicas del modelo, y probar y recalibrar los parámetros del modelo en conjuntos de datos representativos actualizados, la actualización del modelo para evitar premisas que hayan quedado obsoletas<sup>124</sup>.

Por cuanto al segundo grupo, la elección de las técnicas de análisis de datos, se insiste en valorar si es necesario usar métodos de inteligencia artificial, pues otros métodos de análisis suelen ser también ade-

<sup>¿</sup>Son los datos relevantes para el objetivo?; ¿Es la agregación de datos en un nivel suficientemente informativo para el propósito?; ¿Se mantienen los datos de forma coherente a lo largo del tiempo?; ¿Se registran los datos de la misma manera?; ¿Hay suficiente variación en los datos? ¿Están determinadas subpoblaciones (minorías) sobrerrepresentadas en los datos, lo que podría provocar un sesgo en el modelo de predicción?; ¿El conjunto de datos utilizado es correcto, oportuno y completo?; ¿Pueden vincularse las fuentes de datos (si se trabaja con varias tablas)?; ¿Se pueden rastrear las mutaciones de los datos (fuente)?; ¿Ha habido un cambio en la base de datos? MINISTERIE VAN JUSTITIE (2021)... ob. cit. Parte II.

Elección de datos de entrenamiento, posibles factores discriminatorios y documentarlos, registrar qué datos de entrenamiento y qué parámetros se han utilizado, cómo afectaría a los resultados una elección diferente de los parámetros, cómo afectaría a los resultados un conjunto de datos de entrenamiento diferente. Id.

<sup>&</sup>lt;sup>123</sup> WHITE HOUSE (2022), *op.cit.*, pp. 23-29, en especial 26 y ss.

<sup>&</sup>lt;sup>124</sup> NIST; SCHWARTZ, R. y otros (2022), *op.cit.*, pp. 20-27.

cuados. En cualquier caso, hay que justificar la elección de la técnica y las hipótesis preconcebidas<sup>125</sup>. Respecto de la metodología, hay que describir los pasos del desarrollo del modelo respecto de los datos, el tratamiento de datos (preparación y limpieza de datos, composición y estructura de las pruebas modelo, marco de datos básico, creación de variables, corrección de errores de datos). El desarrollo del modelo y la evaluación del modelo. También, por cuanto al uso del algoritmo en la práctica, hay que tener en cuenta el riesgo de que los algoritmos de autoaprendizaje hagan uso de correlaciones y relaciones estadísticas sobre las que es cuestionable que sean adecuadas para la toma de decisiones. Respecto del aprendizaje automático (supervisado), se apunta que hay que hacer explícitos (cuantitativamente) los requisitos relativos a la aplicabilidad, la previsibilidad, la explicabilidad, la discriminación indeseable, etc., de modo que se reduzca aún más el espacio de soluciones. En concreto, se señala que hay que indicar a qué población y las circunstancias en las que debe aplicarse el modelo. Hay que fijar por expertos cómo se evaluará el poder predictivo del modelo, con qué métricas y valores de umbral; o cuándo se determinará si el modelo es suficiente, si falsos positivos son menos malos que los falsos negativos. Así, cabe fijar la discriminación o sesgo indeseable, que sea medible y objetivo. Asimismo, se debe cuantificar el sesgo deseado si lo hav.

En el ámbito de la evaluación y minimización del sesgo y la discriminación ya en 2017 la resolución del Parlamento UE sobre macrodatos "insta" a "minimizar la discriminación y el sesgo algorítmicos" (n° 20) y afirma también la necesaria "mitigación algorítmica" (n° 21, ver también 32). Subraya la posibilidad de corrección de datos y de recurrir decisiones algorítmicas. Los propios algoritmos pueden ser el antídoto contra el sesgo algorítmico y se pueden programar para ignorar o minimizar la importancia que los factores prohibidos en sus decisiones.

AI Watch respecto del sector público señala que los riesgos deben evaluarse sistemáticamente con un procedimiento estructurado y bien definido, se deben identificar las medidas de mitigación adecuadas y ello "debe convertirse en una rutina para las organizaciones

<sup>&</sup>lt;sup>125</sup> MINISTERIE VAN JUSTITIE (2021), *ob. cit.* pp. 23-24.

públicas." <sup>126</sup>. Hay que probar los sistemas para evaluar si se producen sesgos y habrá que documentar los y en su caso ver si es aceptable. Hay que tener en cuenta los cambios en el sistema después del despliegue o de cambios en el contexto de uso o en los datos asociados. La evaluación debe realizarse periódicamente y más los sistemas de mayor riesgo, y siempre que se produzca un patrón de resultados inusuales<sup>127</sup>.

Cuando se detecta la disparidad o sesgo, hay que mitigarlo o eliminarlo. Si no se puede mitigar, debe reconsiderarse el uso del sistema automatizado. Sin embargo, hay que garantizar que el sistema inteligente no pierda su eficacia, su propia naturaleza o lleve a generar resultados absurdos por la eliminación de patrones o datos. No siempre la eliminación es la mejor solución. Por otra parte, resulta prácticamente imposible eliminar todos aquellos datos que revelen la pertenencia a un grupo especialmente protegido sin hacer que el algoritmo sea inservible<sup>128</sup>.

Deben realizarse evaluaciones independientes. Y especialmente se señala que en el caso de los usos del sector público las evaluaciones deberían hacerse públicas salvo restricciones legales o por seguridad nacional. Asimismo, deben presentarse informes incluyendo quién evalúa el sistema y las medidas correctivas que se derivan. Estos informes han de hacer referencia a las evaluaciones de equidad en la fase de diseño, los diseños y pruebas de accesibilidad, las pruebas de disparidad, documentar cualquier disparidad restante y detallar cualquier aplicación y evaluación de mitigación<sup>129</sup>.

El tercer gran bloque es relativo a los factores humanos, incluidos los sesgos sociales e históricos, respecto de los que desde NIST, Unesco y otros ámbitos se establecen orientaciones de actuación bá-

<sup>126</sup> JOINT RESEARCH CENTRE(2022), ob. cit. p. 6.

<sup>&</sup>lt;sup>127</sup> WHITE HOUSE (2022), op.cit., p. 24.

BID- Universidad Adolfo Ibáñez (2022), op.cit., pp. 6-9. También, GOODMAN, B. W. (2016): "A step towards accountable algorithms?: Algorithmic discrimination and the European Union general data protection", 29th Conference on Neural Information Processing Systems, http://www.mlandthelaw.org/ sigo en este caso por SORIANO ARNANZ, A. (2021 b), ob. cit.

<sup>&</sup>lt;sup>129</sup> WHITE HOUSE (2022), op. cit., p. 28.

sicas y se contienen numerosas propuestas. <sup>130</sup> Se destacan como remedios la necesidad de evaluaciones de impacto, la integración en el diseño, desarrollo, implementación y evaluación de las múltiples partes interesadas, priorizar la diversidad, la equidad y la inclusión en los equipos involucrados en la capacitación y el despliegue de sistemas de IA. Además, el NIST analiza las cuestiones de la necesaria gobernanza y supervisión a través de canales de reclamación, políticas y procedimientos, obligaciones de documentación, fijación de responsabilidades y otras recomendaciones<sup>131</sup>.

Especialmente por cuanto a la transversalidad y la integración de los equipos humanos desde Unesco y el Consejo de Europa se insiste en la necesidad de concienciación especial en las profesiones STEM, así como la necesidad de insertar en equipos el enfoque de género 132. Como recuerda Valle Escolano, se necesitan equipos de trabajo diversos y multidisciplinares, se ha de insistir en una cultura de la ética y la responsabilidad en las con controles y equilibrios 133. Si es necesario deben crearse nuevos conjuntos de datos, así como adaptar los existentes. Hay que concienciar y sensibilizar de los problemas que se generan por la falta de representatividad y sesgo, para que así se incremente la recopilación y el uso de datos interseccionales, especialmente de grupos subrepresentados. Se afirma también que hay que comprometerse a usar conjuntos de datos que hayan sido desarrollados con un prisma de igualdad de género (pese a que ello pudiera suponer un aumento de costes).

NIST; SCHWARTZ, R. y otros (2022), op.cit., pp. 35-42; VALLE ESCOLANO, R. (2022), ob. cit.; UNESCO, Artificial Intelligence and Gender Equality, cit.

NIST; SCHWARTZ, R. v otros (2022 a), *op.cit.*, pp. 42-47.

Sobre este tema y necesidad de concienciación especial en profesión, CONSEJO DE EUROPA (2018), op.cit., pp. 55 y ss.; UNESCO (2020): Artificial Intelligence and Gender Equality. Key findings of UNESCO's Global Dialogue, pp. 19 y ss., 22-23, así como la posibilidad de paneles éticos, p. 24, https://unesdoc.unesco.org/ark:/48223/pf0000374174. Para el ámbito laboral, EUROPEAN INSTITUTE OF GENDER EQUALITY (EIGE), Artificial intelligence, platform work and gender equality, 2021, https://eige.europa.eu/publications/artificial-intelligence-platform-work-and-gender-equality-report

VALLE ESCOLANO, R. (2022), ob. cit., p. 99-100.

#### 11.- LA (CRECIENTE) OBLIGACIÓN DE DISEÑO Y ESTUDIOS DE IMPACTO FRENTE A LA DISCRIMINACIÓN ALGORÍTMICA DE LOS SISTEMAS DE IA PÚBLICOS

### 11.1.- Los deberes de responsabilidad proactiva, privacidad en el diseño o estudio de impacto va son exigibles en general a los usos públicos de la IA

Como he tenido ocasión de exponer (Cotino, 2022) ante la IA hay que asumir técnicas de responsabilidad "proactiva, "demostrada" o "accountability" en el diseño, bajo el modelo jurídico del "más vale prevenir que curar" 134. A modo de las leyes de la robótica de Asimov, supone de la inserción del cumplimiento ético y legal en el mismo "Código", en los conocidos términos de Lessig. Se trata de la filosofía del compliance o cumplimiento normativo, de corte anglosajón y en buena medida ajena al Derecho continental europeo. No obstante, es una técnica ya conocida el sector público en ámbitos como el medio ambiente o los ya asentados estudios de impacto de género para disposiciones generales<sup>135</sup>, que sin duda habrán de incorporar la discriminación algorítmica pública. En todo caso, especialmente este sistema de cumplimiento ha sido desarrollado en el contexto de la

135 Impuestos a nivel estatal por artículos 22.2 para leyes y 24.1.b para reglamentos, Ley 50/1997, de 27 de noviembre, del Gobierno; artículo 19 Ley Orgánica 3/2007 para disposiciones de carácter general y los planes de especial relevancia económica y artículos 26. 3 f) Ley 40/2015. Existe abundante normativa auto-

nómica.

Sobre la aplicación del principio de responsabilidad proactiva y el diseño para el ámbito de la IA y el big data mi estudio (2022 a). Destaca, HERNÁNDEZ PE-ÑA, J. C. (2022): El marco jurídico de la inteligencia artificial. Principios, procedimientos y estructuras de gobernanza, Aranzadi, Cízur, pp. 119-172 y también con especial atención a la gobernanza (pp. 173-206). Cabe seguir: ALBERTO GONZÁLEZ, P. (2017): "Responsabilidad proactiva en los tratamientos masivos de datos", Dilemata, nº 24, pp. 115-129; NAVAS NAVARRO, S. (2017): "Derecho e inteligencia artificial desde el diseño. Aproximaciones", NAVAS NAVARRO, S. (coord..). Inteligencia artificial: tecnología, derecho, Tirant lo Blanch, Valencia, pp. 23-72; MARTÍNEZ MARTÍNEZ, R. (2019): "Inteligencia artificial desde el diseño. Retos y estrategias para el cumplimiento normativo", Revista catalana de dret públic, nº 58, pp. 64-81.

protección de datos y, de ahí, se proyecta para los sistemas IA en aplicación de dicha normativa o en la futura regulación del RIA. Precisamente el "Ethics & Rule of law by design X-by design" es la enseña de la "IA confiable" "Made in Europe" 136 que está en la base de las políticas y normativas de IA en la UE. Pues bien, su adopción ha de ser también la viga maestra de la garantía de la evitación de los sesgos, la no discriminación y, en general, el cumplimiento normativo del uso de la IA en el sector público.

En el ámbito de la protección de datos el principio de la responsabilidad proactiva (art. 5, Considerando 78) se traduce en muchas obligaciones<sup>137</sup>, como el registro interno de las actividades del tratamiento, identificando el análisis del riesgo en cada tratamiento (art. 30, art. 31 Ley Orgánica 3/2018, de 5 de diciembre, en conexión con el inventario de actividades en el sector público, art. 77.1°). En todo caso, hay que destacar especialmente la obligación de la protección de datos desde el diseño y por defecto (art. 25) antes del inicio del tratamiento de datos y, singularmente, las evaluaciones de impacto (art. 35 RGPD), esto es, el análisis y descripción de todas las operaciones, su necesidad y la proporcionalidad y la evaluación de los riesgos<sup>138</sup>.

El estudio de impacto será obligatorio en casi todo supuesto de uso público de IA que suponga un tratamiento de derecho de protección de datos siguiendo los criterios de la AEPD de 2019<sup>139</sup>.

HLEG (2019), *op.cit.*, en versión 2018, pp. 19, 21, 29.

d como la diligente elección de un encargado, las garantías suficientes para aplicar medidas técnicas y organizativas apropiadas (art. 28) y la necesidad de decidir cuáles son las adecuadas según los riesgos (art. 32 RGPD), la designación de un delegado de protección de datos (art. 36), la notificación de una violación de la seguridad a las autoridades (art. 33) y su comunicación a los interesados (art. 34).

GRUPO DEL ARTÍCULO 29 (2018), ob. cit. pp. 3 y ss.

Afirma que se han de dar al menos dos de estos supuestos: toma de decisiones automatizadas; perfilados de comportamiento; uso de datos a gran escala; asociación, combinación o enlace de registros de bases de datos para varias finalidades o utilización de nuevas tecnologías o un uso innovador de tecnologías consolidadas, incluyendo la utilización de tecnologías a una nueva escala, con un nuevo objetivo o combinadas con otras, AEPD (2019): Listas de tipos de tratamientos de datos que requieren EIPD (art 35.4), https://www.aepd.es/media/criterios/listas-dpia-es-35-4.pdf

Sobre el cumplimiento de estas medidas cabe remitir a los instrumentos desarrollados por las autoridades europeas<sup>140</sup> y las autoridades de datos españolas, también para el ámbito concreto del sector público<sup>141</sup>. De hecho, en buena medida hay que remitir al conjunto de remedios que se ha hecho referencia que en muy buena medida pueden introducirse en los análisis de riesgos y estudios de impacto como medidas a adoptar para prevenir los riesgos.

A todo este arsenal de medidas preventivas, tecnológicas y de garantías habrá que añadir las especiales garantías de transparencia, explicabilidad y recurribilidad especialmente reforzadas en el caso de las decisiones automatizadas (arts. 22 y 13 RGPD). Además, hay que subrayar que con carácter general todas estas exigencias se harán más intensas en los supuestos de tratamiento de datos especialmente protegidos del artículo 9 RGPD. Como se ha dicho, ello se traducirá en un claro fortalecimiento de garantías en muchos de los supuestos de discriminaciones especialmente protegidas.

Hay que llamar la atención de que, a partir de la actual normativa y principios de la protección de datos, las autoridades de protección de datos ya han interpretado y *extraído* de forma expansiva muchas obligaciones sobre gobernanza de los datos, calidad y robustez de los sistemas de IA, evitación de sesgos y discriminación, obligación de auditorías, así como importantes obligaciones de transparencia. Y todo ello sin esperar a la aprobación de la normativa de IA. Baste seguir las

Así, GRUPO DEL ARTÍCULO 29 (2017): Directrices sobre la evaluación del impacto de la protección de datos (DPIA) y determinar si el procesamiento es "probable que genere un alto riesgo" a los efectos del Reglamento 2016/679, WP248 rev.01, Revisadas por última vez y adoptadas el 4 de octubre de 2017, https://www.aepd.es/sites/default/files/2019-09/wp248rev01-es.pdf

AÉPD (2021 b): Gestión del riesgo y evaluación de impacto en tratamientos de datos personales, https://www.aepd.es/es/node/46578; AEPD (2019): Modelo de informe de Evaluación de Impacto en la Protección de Datos (EIPD) para Administraciones Públicas, https://www.aepd.es/es/node/816; AUTORIDAD CATALANA DE PROTECCIÓN DE DATOS (2017): Guía práctica sobre la evaluación de impacto relativa a la protección de datos, https://apdcat.gencat.cat/es/documentacio/guies\_basiques/Guies-apdcat/Guia-sobre-la-evaluacion-de-impacto-relativa-a-la-proteccion-de-datos-en-el-RGPD/; AEPD (2018): Guía práctica para las evaluaciones de impacto en la protección de datos personales, https://www.lopdencastellon.com/wp-content/uploads/2018/02/Gu%C3%ADa-Evaluaci%C3%B3n-de-impacto-protecci%C3%B3n-de-datos.pdf

Guías sobre IA de la AEPD (2020 y 2021) <sup>142</sup> o del Grupo del artículo 29 (2018) y ahora del CEPD.

Pues bien, las medidas proactivas de protección de datos y sus proyecciones para la IA serán obligatorias y exigibles en tanto en cuanto el uso público de un sistema de IA suponga un tratamiento de datos personales, que como se expuso, no será nada extraño.

### 11.2.- Los usos públicos de IA de alto riesgo habrán de cumplir, además, los deberes del RIA

El enfoque preventivo del cumplimiento de obligaciones según el riesgo es la base del régimen de la propuesta de Reglamento de IA de 2021<sup>143</sup>. Y los sistemas de alto riesgo públicos que se han identificado supra tienen que cumplir todo el grueso de la regulación preventiva y de gobernanza<sup>144</sup>. Tiene especial relación con la evitación de sesgos y discriminación las obligaciones de evaluación y mitigación de riesgos (art. 9 RIA), así como en especial las obligaciones de utilizar conjuntos de datos de alta calidad con buena gobernanza (art. 10 y 16 RIA). Se parte de que los datos "deben estar tan libres de errores y ser tan completos como sea posible" (Considerando 44). Se regula la concreta obligación de hacer un "examen a la vista de posibles sesgos que puedan afectar a la salud y la seguridad de las personas físicas o dar lugar a discriminaciones prohibidas" (art. 10. 2° f). También en el RIA hay conjuntos de obligaciones de los desarrolladores de generar documentación técnica para los clientes o usuarios del sistema (art. 11), el diseño de la IA para que genere registros y logs de funcionamiento (art. 12). Estas obligaciones, además de demostrar el cumplimiento de los deberes, pueden ser decisivas para la comprobación ex post de los

AEPD (2020 a): Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. https://www.aepd.es/es/media/guias/adecuacion-rgpd-ia.pdf, en especial, pp. 30-48 y, con más claridad en AEPD (2021 a)

Un primer análisis de este en España puede seguirse en COTINO HUESO, L. y otros (2021): "Un análisis crítico constructivo de la Propuesta de Reglamento de la Unión Europea por el que se establecen normas armonizadas sobre la Inteligencia Artificial (Artificial Intelligence Act)", Diario La Ley, 2 de julio, Wolters Kluwer. https://links.uv.es/2FK3xc4

En especial, HERNÁNDEZ PEÑA, J. C. (2022), ob. cit.

mismos. Las medidas proactivas frente a la discriminación y el sesgo pasan también en general por las obligaciones de supervisión humana (art. 14) y evitación del "sesgo de automatización". Y especialmente los sistemas de IA públicos evitarán los sesgos y la discriminación si están diseñados con un nivel adecuado de precisión, solidez y ciberseguridad (art. 15). La "solidez" técnica implica obligaciones para prevenir o minimizar "errores, fallos, incoherencias, situaciones inesperadas" (Considerando 50) y la eliminación o reducción de resultados sesgados una vez comercializados (art. 15. 3°). Los sistemas habrán de seguir procedimientos de evaluación de la conformidad (art. 19) y una vez en funcionamiento se someten al sistema de vigilancia del mercado y de supervisión. Todas estas exigencias serán desarrolladas además por todo un armazón de normas técnicas y de armonización, que serán acompañadas de la normalización técnica de la IA que está en fase de desarrollo.

Previsiblemente dos tercios de los sistemas de IA públicos no serán de alto riesgo<sup>145</sup>, por lo no estarán obligados a adoptar estas medidas preventivas. Ahora bien, estas medidas serán los referentes de las buenas prácticas a seguir voluntariamente en códigos, sellos o certificaciones a las que se sumen (Considerando 81, art. 69) los proveedores públicos o privados de esos sistemas públicos de IA. De igual modo, estas medidas podrán integrarse en las que hay que adoptar por protección de datos o en razón de las medidas que se exijan por no discriminación, por ejemplo, en los estudios de impacto de género.

Pues bien, el cumplimento de todas estas obligaciones de responsabilidad proactiva acaba secuenciándose en toda una serie de medidas a adoptar en las diversas etapas de interés: al momento de concebir y diseñar el sistema de IA público, o al momento de preparar los pliegos y mecanismos de contratación de este. También durante el desarrollo o la adquisición del sistema y antes de su despliegue, momento en el que el entrenamiento del sistema, la comprobación de errores y sesgos o la seguridad de este es especialmente relevante. Tras el despliegue del sistema cobra especial protagonismo garantizar la intervención humana y la supervisión de los resultados y la reevaluación de su fun-

Ver supra, JOINT RESEARCH CENTRE (2022), ob. cit. p. 58 Cabe recordar que también en principio, sólo un 10% de sistemas de IA serán de alto riesgo.

cionamiento, así como la garantía de los derechos de los interesados afectados de acceso de información y frente a posibles vulneraciones de derechos<sup>146</sup>.

# 11.3.- La extensión de estudios de impacto y la obligatoriedad de medidas para la no discriminación de los sistemas de IA públicos

La FRA fue de las primeras instituciones en prestar especial atención al fenómeno de la discriminación algorítmica. En su estudio de 2021, su dictamen nº 2 invita a que el legislador de la Unión haga obligatorias las evaluaciones de impacto que cubran toda la variedad de derechos fundamentales<sup>147</sup>, en un sentido similar, el European Law Institute<sup>148</sup>. Como en otros lugares he señalado<sup>149</sup>, siguiendo especialmente a Mantelero<sup>150</sup> el estudio de impacto no ha de ser exclusivo de la protección de datos, sino a todos los derechos digitales y, por supuesto, para el caso de la igualdad, no discriminación y evitación del sesgo. La necesidad de implantar instrumentos preventivos y del estudio de impacto para la igualdad se percibe con mucha claridad en

Un buen ejemplo a seguir lo brinda en especial CEPD (2022): Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement, Version 1.0, 12 mayo, https://edpb.europa.eu/our-work-tools/documents/public-consultations/2022/guidelines-052022-use-facial-recognition\_en Ver, Anexo II, pp. 29-35.

Ver todas las referencias a los diversos trabajos. En particular ahora FRA(2021): Getting the future right artificial intelligence and fundamental rights, FRA, https://fra.europa.eu/en/publication/2020/artificial-intelligence-and-fundamental-rights

EUROPEAN LAW INSTITUTE (2022): Model Rules on Impact Assessment of Algorithmic Decision-Making Systems Used by Public Administration, European Law Institute-University of Viena, https://www.europeanlawinstitute.eu/fileadmin/user\_upload/p\_eli/Publications/ELI\_Model\_Rules\_on\_Impact\_Assessment\_of\_ADMSs\_Used\_by\_Public\_Administration.pdf

<sup>149</sup> COTINO HUESO, L. (2022 a), ob. cit.

Son muchas las obras del máximo nivel del autor desde hace años, puede seguirse en su reciente monografía MANTELERO, A. (2022): Beyond Data. Human Rights, Ethical and Social Impact Assessment, Springer, Information Technology and Law Series IT&LAW 36, 2022, https://link.springer.com/book/10.1007/978-94-6265-531-7

Recomendación Unesco de 2021<sup>151</sup>, especialmente en el capítulo dedicado a la "Equidad y no discriminación" (n° 28 y ss.)<sup>152</sup>; en el ámbito de actuación 1° de "evaluación del impacto ético" (n° 50-53)<sup>153</sup> y el ámbito de actuación 6 de "género" (n° 87-93) el n°87<sup>154</sup>.

De modo más concreto, la evitación de sesgos y de discriminación se integran en los marcos de gestión de riesgos que para la IA se están estandarizando, por ejemplo, desde el NIST en 2022<sup>155</sup>. Para el sector público cabe seguir grandes avances como en Países Bajos, ya referenciados y especialmente incisivos respecto de la no discriminación. Hay que estimular que en España se sigan estas buenas prácticas.

La misma AEPD respecto de los estudios de impacto de protección de datos (que serán por lo general obligatorios en los usos de IA pública) recuerda la necesidad de atender a otros derechos, así como específicamente a la no discriminación. Pues bien, como se expuso, los sistemas de IA públicos que hagan diferencias de trato por circunstancias especialmente sospechosas, por lo general supondrán tratamientos de datos especialmente protegidos. Y en estos supuestos los estudios de impacto han de ser especialmente incisivos en la detección de riesgos y las medidas compensatorias frente a los mismos,

Así, todos los derechos que se han de tener en cuenta "a lo largo del ciclo de vida de los sistemas de IA" (nº 16) y en particular "la protección y la promoción de la diversidad y la inclusión" con especial "participación activa de todas las personas o grupos" (nº 19).

Los actores de la IA deberían hacer todo lo razonablemente posible por reducir al mínimo y evitar reforzar o perpetuar aplicaciones y resultados discriminatorios o sesgados a lo largo del ciclo de vida de los sistemas de IA, a fin de garantizar la equidad de dichos sistemas." (n°29), también "abordar las brechas digital y de conocimientos" (n° 30).

<sup>&</sup>quot;Los gobiernos deberían adoptar un marco regulador que establezca un procedimiento para que las autoridades públicas, en particular, lleven a cabo evaluaciones del impacto ético de los sistemas de IA [...] Se debería exigir a las autoridades públicas que supervisen los sistemas de IA que hayan implantado o desplegado, mediante la introducción de mecanismos e instrumentos adecuados." (nº 53).

El n°87 insiste en que la evaluación del impacto ético debería incluir una perspectiva transversal de género. Y que los estereotipos de género y los sesgos discriminatorios no se trasladen a los sistemas de IA, sino que se detecten y corrijan de manera proactiva." (n° 90).

NIST (2022): AI Risk Management Framework: Initial Draft, 17 de marzo, https://www.nist.gov/document/ai-risk-management-framework-initial-draft

habiendo de integrar en la medida de lo posible las buenas prácticas y recomendaciones habitualmente señaladas. Es más, en estos supuestos, se hace precisa una regulación legal y con garantías para que la Administración efectúe estos tratamientos de datos. Por ello, los instrumentos legales habrán de integrar y concretar las garantías para evitar el sesgo, error y discriminación en los sistemas públicos de IA. De este modo, no se tratará de la adopción voluntaria de buenas prácticas y recomendaciones.

Y ahí es donde han de cobrar especial importancia los estudios de impacto de género obligatorios para disposiciones generales, conectados con los muchos supuestos en los que las diferencias de trato y los sistemas de IA deben ser regulados por ley o reglamento. Hoy por hoy, las herramientas y guías para la realización de estos estudios de impacto normativo<sup>156</sup> no incluyen las cuestiones que aquí interesan, pero sin duda que se van a ir generalizando e incluyendo estos problemas debido a la evolución y los cambios normativos recientes.

En España, además de las actuales exigencias por la protección de datos, es del todo acertada la Carta de Derechos Digitales que afirma en general "el principio de cumplimiento normativo desde el diseño" para todos los ámbitos en su apartado I. 4 <sup>157</sup>. Ya para la IA hace expresa mención a "garantizar el derecho a la no discriminación

Entre otros, cabe seguir, DELGADO GOBOY y otras (2003): Guía de aplicación práctica, para la elaboración de informes de impacto de género de las disposiciones normativas que elabore el gobierno, de acuerdo a la ley 30/2003, 2003, Instituto De La Mujer (Ministerio de Trabajo Y Asuntos Sociales), https://www.inmujeres.gob.es/publicacioneselectronicas/documentacion/Documentos/DE0259.pdf

LIKADI (2018): Guía para la evaluación de impacto de género de la normativa guía para la evaluación de impacto de género de la normativa, Diputación Foral de Álava, 2018, https://web.araba.eus/documents/105044/1393927/Gu%C3%ADapara+la+evaluaci%C3%B3n+de+impacto+de+g%C3%A9nero+%C3%81lava+castellano+navegable.pdf/74905ae0-fc50-7b15-a5c3-2ab4b2f16fa3?t=1557493191803

Pueden seguirse otras en https://presidencia.gva.es/es/guias-para-elaborar-informes-de-impacto-de-genero

<sup>&</sup>quot;declara que el principio de cumplimiento normativo desde el diseño es de aplicación íntegramente al desarrollo de los entornos digitales y por ello los desarrollos científicos, tecnológicos y su despliegue contemplarán en la determinación de sus requerimientos un análisis sobre el cumplimiento de tal principio."

cualquiera que fuera su origen, causa o naturaleza, en relación con las decisiones, uso de datos y procesos basados en inteligencia artificial" (apartado V). Y para el sector público se hace referencia a la "no discriminación, en particular por razón de sexo, de las tecnologías usadas por las Administraciones públicas" (XVIII. 3°) y, por lo que más interesa, que "Será necesaria una evaluación de impacto en los derechos digitales en el diseño de los algoritmos en el caso de adopción de decisiones automatizadas o semiautomatizadas." (XVIII. 7°). Además, las evaluaciones han de alcanzar las leyes administrativas (apartado VII. 4°).

Como primera proyección el artículo 25. 1º Ley 15/2022, de 12 de julio afirma que "1. La protección frente a la discriminación obliga a la aplicación de métodos o instrumentos suficientes para su detección, la adopción de medidas preventivas". Y el nuevo artículo 23 Ley 15/2022, de 12 de julio es el mejor exponente cuando afirma "los algoritmos involucrados en la toma de decisiones que se utilicen en las administraciones públicas tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas, siempre que sea factible técnicamente. En estos mecanismos se incluirán su diseño y datos de entrenamiento, y abordarán su potencial impacto discriminatorio. Para lograr este fin, se promoverá la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio" (1º) y "priorizarán la transparencia en el diseño y la implementación y la capacidad de interpretación de las decisiones adoptadas por los mismos." (2º).

La nueva ley emplea términos muy laxos ("favorecerán", "tengan en cuenta", "siempre que sea factible", "se promoverá", "priorizarán", "promoverán"). Sin embargo, es muy posible que la necesidad de regular el uso de IA público que aquí se sostiene dé protagonismo práctico a los estudios de impacto que sí que son obligatorios. Dado que estos estudios de impacto normativo de género en el futuro van a ser sensibles e irán incluyendo estas cuestiones, especialmente para el sector público, es muy posible que generen un importante efecto preventivo de la discriminación algorítmica en la práctica. Bien es cierto que no se trataría estrictamente de una obligación hasta que haya una mejora regulatoria o hasta que las autoridades regulatorias competentes o judiciales hagan una interpretación más activista de las normas.

En todo caso, siempre *quedará el París* de la obligatoriedad de medidas proactivas por la protección de datos, con su interpretación claramente expansiva y proyectable para el sector público. Y también, todos los deberes del RIA para los usos públicos de la IA de alto riesgo.

### 11.4.- La necesidad de una visión social y colectiva y de integrar garantías de calidad de datos y de participación

Las medidas proactivas y preventivas no sólo deben tener un enfoque en el individuo, sino en la prevención de la igualdad y la no discriminación desde el interés de los colectivos, la sociedad y el sistema democrático en su conjunto. No olvidemos el impacto estructural y masivo del uso de estos sistemas. Además, muchos sistemas IA públicos se utilizan para apoyar la toma de decisiones generales políticas o normativas. En estos supuestos, aunque los sistemas no manejen datos personales, también habrán de darse exigencias de los principios de calidad y buena gobernanza y gestión respecto de los datos no personales. Asimismo, habrá de asegurarse la particular transparencia y la participación de la sociedad civil en estos contextos de uso. Ello puede articularse en el marco de procedimientos participativos, de audiencia pública que se adopten o su inserción en órganos colegiados. Así lo propuse para la Carta de Derechos Digitales en un apartado sobre "Garantías de impacto social"158 y de nuevo resulta especialmente inspirador Mantelero<sup>159</sup>. La Recomendación Unesco es llamativamente insistente en la necesidad de un enfoque colectivo y sobre todo de la inclusión de mecanismos de participación social respecto del uso de sistemas IA160. Para la IA pública, también el European Law

Ver mi aportación en RED DAIA (2020 b): Aportación DAIA a consulta pública Carta de derechos digitales, diciembre 2020 https://bit.ly/3paF0H0

Además de la obra reseñada, por ejemplo, (2017): "From group privacy to collective privacy: towards a new dimension of privacy and data protection in the big data era", en TAYLOR. L.; VAN DER SLOOT, B.; FLORIDI, L. (eds.) (2017), ob. cit. cap. 8.

Así en el apartado específico de "Equidad y no discriminación" se menciona la "participación inclusiva de colectivos" (n° 28) y la "participación significativa" de la sociedad en la evitación de brechas a lo largo del ciclo de vida de la IA (n°

Institute hace interesantes propuestas de participación en estudios de impacto (art. 6. 2 c) ii), auditorías (art. 10) y participación pública (art. 11).

Recuerda Gutiérrez las Recomendaciones de la APDCAT de publicar la EIPD relativas a los tratamientos que son responsabilidad de las Administraciones públicas, pues "resulta de especial relevancia cuando el interés del tratamiento puede llegar a suponer un cierto impacto social o puede generar un estado de opinión en las personas los datos de las cuales pueden llegar a ser objeto de tratamiento" 161.

# 12.- Y SI LOS REMEDIOS NO VALEN, ¿QUÉ PASA CUANDO SE DETECTA UNA DISCRIMINACIÓN ALGORÍTMICA PÚBLICA?

## 12.1.- Obligaciones de mitigación, rectificación, corrección e incluso cese de uso del sistema de IA público

Según se ha expuesto, de modo proactivo y con carácter general los proveedores o usuarios de sistemas de IA públicos han de evaluarlos con pruebas y mitigación de sesgos de forma continua y con supervisión organizativa. Las evaluaciones en su caso han de ser independientes. De modo reactivo, si se detecta un sistema de IA público sesgado o discriminatorio, se deben adoptar medidas de mitigación, modulación, corrección o eliminación de las causas. En su caso, habrá que bloquear o cesar el uso del sistema. Como ahora se sigue, estas obligaciones son claras desde la normativa de igualdad, de procedimiento Administrativo, de protección de datos y el RIA.

161

<sup>30).</sup> En los números 46 y 47 sobre "Gobernanza y colaboración adaptativas y de múltiples partes" se incide en que "la participación de las diferentes partes interesadas a lo largo del ciclo de vida de los sistemas de IA es necesaria para garantizar enfoques inclusivos de la gobernanza de la IA". Y la participación ciudadana debe facilitarse respecto de la evaluación del impacto ético (n° 50, 52, 53), lo mismo que se reitera en los números 60 y 64 o en el n° 91 respecto la participación en el ámbito de actuación 6 de género.

Desde la normativa de igualdad, cabe recordar que el artículo 25. 1º Ley 15/2022 además de las medidas preventivas y de detección, "obliga" a "la articulación de medidas adecuadas para el cese de las situaciones discriminatorias", asimismo, "3. Ante un incidente discriminatorio, las autoridades encargadas de hacer cumplir esta ley tomarán las medidas oportunas para garantizar que los hechos no vuelvan a repetirse, especialmente en los casos en los que el agente discriminador sea una administración pública." En el mismo sentido, se han de "adoptar las medidas oportunas y proporcionadas para su eliminación" (art. 31.1°). En razón del artículo 31 habrá de incoarse el procedimiento Administrativo correspondiente y la investigación. Deben recordarse además las particularidades que puedan darse en sectores públicos específicos regulados en los artículos 9 a 24, como en la educación (art. 13), atención sanitaria (art. 15), en la prestación de los servicios sociales (art. 16), en la oferta al público de bienes y servicios (art. 17), en el ámbito de la seguridad ciudadana (art. 18), en la administración de justicia (art. 19), o en culturales y deportivas (art. 23).

En el ámbito del procedimiento, si se considera que el sesgo del sistema de IA no tiene suficiente relevancia, la detección y corrección del sistema y de sus consecuencias podría articularse a través del artículo 109. 2º LPAC. Así, podría considerarse "rectificar en cualquier momento, de oficio o a instancia de los interesados, los errores materiales, de hecho o aritméticos existentes en sus actos". También, ante la detección o denuncia de un sesgo o discriminación algorítmica en el procedimiento, incluso antes del inicio formal del mismo, no será extraña la adopción de medidas provisionales (art. 56 LPAC) e incluso la posibilidad de la suspensión de actos (art. 108 LPAC). La denuncia por el afectado o interesado del sistema de IA podrá implicar la suspensión de la ejecución del acto o actuación impugnada, por tratarse de una posible causa de nulidad de pleno por discriminación, bien por los perjuicios de imposible o difícil reparación (art. 117. 2º LPAC).

En el ámbito de protección de datos, los posibles incumplimientos bien pueden estar relacionados con la discriminación o sesgo algorítmicos. El responsable y el encargado del tratamiento tienen deberes de asegurar el cumplimiento de las obligaciones que les corresponden y de actuar en caso de que se detecten deficiencias. Y, obviamente, ante estas situaciones, denuncias de terceros o por la actuación misma de las autoridades de control, deberán cumplir las órdenes de las

autoridades (art. 58 RGPD). Ello puede suponer la paralización del sistema de IA público: las "medidas provisionales y de garantía de los derechos" pueden incluir el "bloqueo de los datos y la cesación de su tratamiento", incluso "su inmovilización" (art. 69 Ley orgánica 3/2018), igualmente la resolución de una infracción incluirá las medidas "para que cese la conducta o se corrijan los efectos de la infracción" (art. 77.2° Ley orgánica 3/2018).

Bajo el RIA, los proveedores o usuarios de sistemas IA públicos de "alto riesgo" están obligados a adoptar "medidas correctoras" en cuanto detecten que el sistema no cumple con el RIA (art. 16); "investigarán inmediatamente, en su caso, las causas" (art. 21). También informarán a los organismos y autoridades competentes (art. 22). La autoridad de vigilancia del mercado que aprecie un riesgo para la igualdad del sistema de alto riesgo tendrá que evaluar el sistema y "exigirá sin demora indebida al agente económico correspondiente que adopte todas las medidas correctoras adecuadas", en su caso, "todas las medidas provisionales adecuadas para prohibir o restringir la comercialización del sistema de IA en su mercado nacional, retirar el producto de ese mercado o recuperarlo". También habrá de informar a las autoridades competentes, incluso advertir a la Comisión y otros Estados si el riesgo va más allá de España (art. 65 RIA).

## 12.2.- Nulidad, anulabilidad e inderogabilidad de los sistemas de IA público y sus efectos

En el caso de una discriminación algorítmica de relevancia, hay que partir de que "son nulos de pleno derecho las disposiciones, actos o cláusulas de los negocios jurídicos que constituyan o causen discriminación" en el caso de las discriminaciones especialmente prohibidas (art. 26, Ley 15/2022, de 12 de julio). En la misma línea y como principio, se trataría de causa de nulidad de pleno derecho del artículo 47. 1. a) LPAC, de 1 de octubre por cuanto los sistemas IA "lesionen los derechos y libertades susceptibles de amparo constitucional" 162.

Sobre el tema, en general. GAMERO, E. y FERNÁNDEZ, S. (2016), ob. cit. pp. 532 y ss. y MELLADO RUIZ, L. (2017), "Supuestos de nulidad de pleno derecho de los actos administrativos", en GAMERO CASADO, E. (dir.), Tratado de Procedimiento Administrativo Común y Régimen Jurídico Básico del sector

En el caso de discriminaciones y sesgos no vinculados con especiales causas de no discriminación, no hay que excluir la lesión del artículo 14 CE y la consiguiente nulidad radical. Según se ha visto, en no pocos supuestos el sistema de IA público ha de venir regulado en leyes o en su caso disposiciones generales. Si la causa de la discriminación algorítmica está establecida normativamente ello provocaría la nulidad de pleno Derecho de la norma (art. 47. 2° LPAC). También puede darse el caso de la nulidad de las resoluciones de adopción del sistema u otros instrumentos en los que se haya formalizado la causa de la discriminación. En los casos en los que las medidas y decisiones técnicas en el diseño, implantación o aplicación del sistema se separen de las normas que lo regulan, se dará la inderogabilidad singular del artículo 37 LPAC.

Por cuanto a los efectos que genere el sistema público de IA discriminatorio, en su caso y dependiendo de diversos factores, puede ser aplicable la anulabilidad del artículo 48 LPAC. Posiblemente se aplicará a supuestos de discriminación no especialmente prohibida, en estos casos por lo general la discriminación convergerá con vicios de arbitrariedad y no motivación, por lo que es posible que se canalicen en mayor medida como anulables. Asimismo, es bastante posible la anulabilidad en el caso de considerarse que hay una vulneración del artículo 18 CE, que rara vez se considera que genere nulidad y como Valero subraya desde siempre, no hay que obviar que el artículo 18. 4º CE no sólo es relativo a la protección de datos sino a otro tipo de amenazas de la "informática" 163.

público Vol. 2, Tirant lo Blanch, Valencia, pp. 1929-1983; JORDANO FRAGA, J. (1996): "Nulidad de los actos administrativos y derechos fundamentales, Marcial Pons, Madrid y en (2011): "La nulidad de los actos que lesionen derechos fundamentales: evolución jurisprudencial y desarrollo de la causa de nulidad, núcleo de las garantías en el Estado de Derechos Fundamentales", en LÓPEZ MENUDO F. (coord.), Derechos y garantías del ciudadano: estudios en homenaje al profesor Alfonso Pérez Moreno, 2011, pp. 715-743.

Se trata de un tema complejo, estudiado especialmente por Valero en diversos de sus estudios como VALERO TORRIJOS, J. (2013): Derecho, innovación y administración electrónica, Global Law = Derecho Global, Sevilla. O, FERNÁNDEZ SALMERÓN M. (2004): La protección de los datos personales en las Administraciones Públicas, Aranzadi, Cízur, pp. 465 a 470.

Es muy posible también que haya de aplicarse el artículo 49 LPAC en los casos en los que la IA pública sólo se utiliza en partes del procedimiento que puedan considerarse independientes. En consecuencia, la "nulidad o anulabilidad de un acto no implicará la de los sucesivos en el procedimiento que sean independientes del primero" (art. 49. 1°). En todo caso, habrá que valorar si la discriminación producida, esto es "la parte viciada sea de tal importancia que sin ella el acto administrativo no hubiera sido dictado" (art. 49. 2°). Además, la detección de un fallo en el sistema de IA público utilizado podría suponer una revisión de disposiciones y actos nulos de oficio o a solicitud de los interesados (art. 106 LPAC) o provocar una revocación de actos (art. 109. 1°) o una revisión de actos nulos del artículo 106.

### 12.3.- Los efectos de la discriminación algorítmica en el contexto de la contratación administrativa

Cabe remitir a esta obra y otros estudios respecto de la contratación por el sector público del desarrollo o uso de sistemas tecnológicos y de IA<sup>164</sup>. En EEUU recientemente se recomiendan los desarrollos *in-house*<sup>165</sup>. En cualquier caso, cabe señalar que la Administración contratante del sistema de IA cuenta con generales facultades de inspección (art. 190 Ley 9/2017, de 8 de noviembre). Habrá que estar al marco del contrato específico y su modalidad, pero si el sistema de IA

FREEMAN ENGSTROM, D. y otros (2020), op.cit., pp. 7, 18, 71 y ss.

165

<sup>164</sup> Además del estudio en esta obra sobre el tema interesa LOZA CORERA, M. (2022): "Explotación y regulación del uso del big data e inteligencia artificial para los servicios públicos y la ciudad inteligente", en COTINO HUESO, L. y TODOLÍ SIGNES, A. (coord.), ob. cit. pp. 125-148. Sobre la contratación de IA policial, JOH, E. E. (2017): "The Undue Influence of Surveillance Technology Companies on Policing" (27 febrero). N.Y.U. L. Review Online, pp. 18-20 https://ssrn.com/abstract=2924620 Muchos de los problemas proyectables para la IA los aborda VALERO TORRIJOS, J. (2015): "Ciudades inteligentes y datos abiertos: implicaciones jurídicas para la proteccion de los datos de carácter personal", Istituzioni del federalismo: rivista di studi giuridici e politici, nº. 4, 2015 ("Smart Cities" e amministrazioni intelligenti), pp. 1025-1047 y VESTRI, G. (2021): "Sistemas de inteligencia artificial en la contratación pública: entre códigos fuente y datos abiertos", Actualidad administrativa, nº 12. CONSEJO DE EUROPA (2018), op.cit., pp. 57-58 insiste en la necesidad de participar desde una fase temprana en la adquisición de sistemas de IA público.

desarrollado genera sesgos y discriminación puede darse un supuesto de incumplimiento parcial o cumplimiento defectuoso (art. 192). En el caso habitual de que se articule a través de un contrato de prestación de servicios, "2. El contratista será responsable de la calidad técnica de los trabajos que desarrolle y de las prestaciones y servicios realizados, así como de las consecuencias que se deduzcan para la Administración o para terceros de las omisiones, errores, métodos inadecuados o conclusiones incorrectas en la ejecución del contrato." Si se considera que no se ajusta a lo que corresponde, se podrá rechazar o reclamar la subsanación de los mismos (art. 311). Si el uso del sistema de IA se inserta en un contrato de concesión de servicios, se puede considerar que del funcionamiento del sistema de IA "se derivase perturbación grave y no reparable por otros medios en el servicio" y, en consecuencia, podría intervenirse o secuestrar el contrato (art. 293).

Que el sistema desarrollado genere sesgos o discriminación puede tener diversas consecuencias, como penalidades, daños y perjuicios al proveedor del sistema (art. 194). La exigencia de responsabilidad podrá incluir a los terceros, como lo serían los afectados por el sistema de IA (art. 196). La detección de sesgos y discriminación en su caso podría suscitar una modificación del contrato (arts. 203 y ss. en general), incluso podría darse la resolución del contrato en algunos casos, como por ejemplo por supresión del servicio por razones de interés público (art. 294 d).

### 12.4.- Responsabilidad por el uso de IA público

La discriminación causa responsabilidad patrimonial y la obligación de reparación del daño y cabe recordar que "Acreditada la discriminación se presumirá la existencia de daño moral" (art. 27 Ley 15/2022, de 12 de julio) y cabe acudir al régimen de la responsabilidad patrimonial de las Administraciones Públicas (arts. 32 y ss. LRJSP)<sup>166</sup>. La normativa propia del sector público muy posiblemente habría de

Cabe remitirse al estudio sobre la responsabilidad en esta obra. La literatura en general ya es muy abundante. No obstante, para el ámbito del sector público, una primera aproximación en ALCOLEA AZCÁRRAGA, C., "La responsabilidad patrimonial de la Administración y el uso de algoritmos", *Revista General de Derecho Administrativo*, n°. 59, 2022.

conjugarse con el régimen más general de responsabilidad por uso de IA. Cabe ahora simplemente recordar que se trata de un tema de extraordinaria complejidad y actualidad, al punto de que el 28 de septiembre de 2022 se ha lanzado la Propuesta de Directiva relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial (Directiva sobre responsabilidad en materia de IA)<sup>167</sup>. La misma es especialmente destacable en materia de carga de la prueba y presunción de responsabilidad, en particular su artículo 4. En el mismos se tienen especialmente en cuenta el cumplimiento del deber de diligencia que sea aplicable, así como si el sistema es o no de alto riesgo en el marco del RIA.

### 13.- ¿QUIÉNES PUEDEN REACCIONAR ANTE UNA DISCRIMINACIÓN ALGORÍTMICA Y QUÉ AUTORIDADES DEBEN ACTUAR?

En el caso de la posible discriminación por un sistema de IA público, por una parte, los interesados y potenciales reclamantes serán los titulares del derecho de igualdad o los derechos y los intereses legítimos afectados por el uso del sistema de IA público (art. 4 LPAC). Ahora bien, por otra parte, lo más normal será que haya una "pluralidad indeterminada o de difícil determinación" de personas afectadas y, en consecuencia, se dará el presupuesto por el que serán interesados amplios sujetos de la sociedad civil, además, en la mayoría de los casos sin necesidad de autorización de los afectados individualmente. Ello es así debido a la previsión expresa del artículo 31. Ley 15/2022, de 12 de julio 168. Y en el contexto de la protección de datos por el

https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:52022PC0496 https://ec.europa.eu/info/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence\_en

<sup>68 &</sup>quot;2. A los efectos de lo establecido en el artículo 4.2 de la LPAC, los sindicatos, las asociaciones profesionales de trabajadores autónomos, las organizaciones de personas consumidoras y usuarias y las asociaciones y organizaciones legalmente constituidas que tengan entre sus fines la defensa y promoción de los derechos humanos y cumplan los requisitos fijados en el artículo 29 de la presente ley, podrán tener la consideración de interesado en los procedimientos administrativos en los que la Administración tenga que pronunciarse en relación con una situación de discriminación prevista en esta ley, siempre que cuenten con la auto-

que se puede vehicular una acción, cabe recordar que además de los interesados son posibles las acciones colectivas en virtud del artículo 80 RGPD. Este precepto ha sido recientemente interpretado por la STJUE de 28 de abril de 2022, asunto C-319/20. Asimismo y en su caso, puede llegar a ser aplicable la Directiva (UE) 2020/1828, de 25 de noviembre que regula también las acciones colectiva y su transposición en España. En los supuestos en los que el sistema de IA público venga regulado en una disposición general, cualquier persona podrá impugnar directamente la misma (art. 26 LJCA).

Por cuanto a las autoridades que puedan actuar potencialmente, son muchas debido al tipo de Administración que use el sistema de IA, la finalidad de uso, efectos y muchos otros factores. En todo caso cabe destacar las autoridades de protección de datos, supervisión de algoritmos y de igualdad con las facultades de actuación cuando se detecta un incumplimiento normativo.

Dada la fácil conexión del uso de IA pública con la aplicación del régimen de protección de datos, será muy fácil que de oficio o por denuncia actúen las autoridades estatal o autonómicas. Asimismo, la Agencia Española de Supervisión de Inteligencia Artificial (AESIA) ya prevista legalmente<sup>169</sup> en principio está pensada para la asunción rápida de todas las obligaciones que imponga del RIA (artículo 59 RIA). Según su regulación definitiva, bien se le puede facultar para asumir competencias más allá de los sistemas de alto riesgo, como muchos de los sistemas de IA públicos.

También, la generación de sesgos y discriminaciones por el sector público será de natural el ámbito de actuación de la Autoridad Independiente para la Igualdad de Trato y la No Discriminación creada por el artículo 40 Ley 15/2022, de 12 de julio, con facultades de

rización de la persona o personas afectadas. No será necesaria esta autorización cuando las personas afectadas sean una pluralidad indeterminada o de difícil determinación, sin perjuicio de que quienes se consideren afectados puedan también participar en el procedimiento."

La Ley 22/2021, de 28 de diciembre, de Presupuestos Generales del Estado para el año 2022 autorizó al Gobierno a crear la AESIA para llevar a "cabo medidas destinadas a la minimización de riesgos significativos sobre la seguridad y salud de las personas, así como sobre sus derechos fundamentales, que puedan derivarse del uso de sistemas de inteligencia artificial".

"c) Iniciar, de oficio o instancia de terceros, investigaciones sobre la existencia de posibles situaciones de discriminación", "d) Ejercitar acciones judiciales", "e) Interesar la actuación de la Administración del Estado para sancionar las acciones u omisiones". Como recuerda el Consejo de Europa, los organismos de igualdad, entre otras posibilidades, deben exigir que el sector público se asegure de que cuenta con suficientes conocimientos jurídicos y técnicos para evaluar y supervisar los riesgos. Asimismo, exigir que se evalúen periódicamente si sus sistemas de IA tienen efectos discriminatorios<sup>170</sup>.

#### **BIBLIOGRAFÍA CITADA**

- ABA CATOIRA, A. (2022), "Derechos de igualdad, personas con discapacidad y mayores en el entorno digital (VIII, XI Y XII)", COTINO HUESO, L. (editor), *La Carta de Derechos Digitales*, Tirant Lo Blanch, Valencia, pp. 123-154.
- AEPD (2017), Informe AEPD 385661/2017 de marzo relativo al anteproyecto de ley de sistema de alertas tempranas de la Comunidad Valenciana. No publicado. He dipuesto un enlace en https://www.dropbox.com/s/utzo5hrkf0b3cmi/aepdantifraude.pdf?dl=0
- AEPD (2018): Guía práctica para las evaluaciones de impacto en la protección de datos personales, https://www.lopdencastellon.com/wp-content/uploads/2018/02/Gu%C3%ADa-Evaluaci%C3%B3n-de-impacto-protecci%C3%B3n-de-datos.pdf
- AEPD (2019): Listas de tipos de tratamientos de datos que requieren EIPD (art 35.4), https://www.aepd.es/media/criterios/listas-dpia-es-35-4.pdf
- AEPD (2019): Modelo de informe de Evaluación de Impacto en la Protección de Datos (EIPD) para Administraciones Públicas, https://www.aepd.es/es/node/816
- AEPD (2020 a): Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. https://www.aepd.es/es/media/guias/adecuacion-rgpd-ia.pdf
- AEPD (2020 b): *Privacidad de grupo*, 19 de octubre, https://www.aepd.es/es/prensa-y-comunicacion/blog/privacidad-de-grupo
- AEPD (2021 a): Requisitos para Auditorías de Tratamientos que incluyan IA, 2021.

<sup>&</sup>lt;sup>170</sup> CONSEJO DE EUROPA (2018), op.cit., p. 57.

- AEPD (2021 b): Gestión del riesgo y evaluación de impacto en tratamientos de datos personales, https://www.aepd.es/es/node/46578
- ALBERTO GONZÁLEZ, P. (2017): "Responsabilidad proactiva en los tratamientos masivos de datos", Dilemata, nº 24, pp. 115-129.
- ALCOLEA AZCÁRRAGA, C. (2022): "La responsabilidad patrimonial de la Administración y el uso de algoritmos", Revista General de Derecho Administrativo, nº. 59.
- ALGEMENE REKENKAMER (Tribunal de Cuentas), (2020): Digitaal Toetsingskader Algoritmes (Marco de evaluación digital de Algoritmos), Tribunal de Cuentas Excel, https://www.rekenkamer.nl/binaries/rekenkamer/documenten/publicaties/2021/01/28/download-het-toetsingskader/Toetsingskader+algoritmes+v1.0.xlsx
- AUTORIDAD CATALANA DE PROTECCIÓN DE DATOS (2017): Guía práctica sobre la evaluación de impacto relativa a la protección de datos, https://apdcat.gencat.cat/es/documentacio/guies\_basiques/Guies-apdcat/Guia-sobre-la-evaluacion-de-impacto-relativa-a-la-proteccion-de-datos-en-el-RGPD/
- BALAGUER CALLEJÓN, M. L. (2018): "Acciones positivas para la igualdad efectiva de mujeres y hombres", VENTURA FRANCH, A. y GARCÍA CAMPÁ. S. (coord.) El derecho a la igualdad efectiva de mujeres y hombres: una evaluación del primer decenio de la Ley Orgánica 3/2007, pp. 285-312.
- BAROCAS, S. y SELBST, A. D. (2016): "Big Data's disparate impact". 104 Calif Law Rev 671, http://dx.doi.org/10.2139/ssrn.2477899
- BARONA, S. (2021): Algoritmización del derecho y de la justicia: De la inteligencia artificial a la Smart Justice, Tirant lo Blanch, Valencia.
- BEN- SHAHAR, O. y PORAT, A. (2021): Personalized law: different rules for different people, Oxford University Press.
- BENÍTEZ PALMA, E. (2022): "Mucho ruido y pocas nueces: algunas consideraciones sobre los registros públicos de algoritmos", VESTRI, G. (2022) *ob. cit.* pp. 83-96.
- BID- Universidad Adolfo Ibáñez (2022): Formulación ética de proyectos de ciencia de datos, Banco Interamericano de Desarrollo, División de Gobierno Digital y Universidad Adolfo Ibáñez, https://goblab.uai.cl/guiaformulacion-etica-de-proyectos-de-ciencia-de-datos/
- BOGEN, M. (2019): "All the Ways Hiring Algorithms Can Introduce Bias". Harvard Business Review. Mayo. https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias
- BOIX PALOP, A. (2020): "Los algoritmos son reglamentos: la necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones", en

Revista de Derecho Público: Teoría y Método, Vol. 1, 2020. https://doi.org/10.37417/RPD/vol\_1\_2020\_33

- CEPD (2022): Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement, Version 1.0, 12 mayo, https://edpb.europa.eu/our-work-tools/documents/public-consultations/2022/guidelines-052022-use-facial-recognition\_en
- CERRILLO I MARTÍNEZ, A. (2022): "La personalización de servicios digitales", en CERRILLO I MARTÍNEZ, A. (coord.) *La Administración Digital*, Dykinson, Madrid, pp. 311-370.
- CERRILLO MARTÍNEZ, A. (2020): "¿Son fiables las decisiones de las Administraciones públicas adoptadas por algoritmos?", European Review of Digital Administration & Law Volume 1 Issue 1-2 2020, DOI: 10.4399/97888255389603, págs. 18-36
- CHRISTIAN, A. (2022): Four Ways the TSA Is Making Flying Easier for Transgender People. American Civil Liberties Union. https://www.aclu.org/news/lgbtq-rights/four-ways-the-tsa-is-making-flying-easier-for-transgender-people
- CITRON, D. K. (2008): "Technological Due Process", 85 Wash. U. L. Rev. pp. 1249-1313, pp. 1254 y ss. p. 1286, http://openscholarship.wustl.edu/law\_lawreview/vol85/iss6/2
- COBREROS MENDAZONA, E. (2997): "Discriminación por indiferenciación: estudio y propuesta", *Revista española de derecho constitucional*, Año nº 27, Nº 81, 2007, págs. 71-114.
- COMISIÓN EUROPEA, RENDA A. (2021): Study to Support an Impact Assessment of Regulatory Requirements for Artificial Intelligence in Europe. Final Report (D5), abril 2021. https://op.europa.eu/es/publication-detail/-/publication/55538b70-a638-11eb-9585-01aa75ed71a1
- CONSEJO DE EUROPA; ZUIDERVEEN BORGESIUS F (2018): Discrimination, artificial intelligence, and algorithmic decision-making. Directorate General of Democracy, Consejo de Europa. https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73
- COTINO HUESO L. y CASTELLANOS CLARAMUNT J. (coords.), *Transparencia y explicabilidad de la inteligencia artificial*, Tirant lo Blanch, Valencia, 2022.
- COTINO HUESO, L. (2017): "Big data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales", en *Dilemata. Revista Internacional de Éticas Aplicadas*, nº 24, 2017. pp. 131-150, https://goo.gl/iERVha
- COTINO HUESO, L. (2020 a): "'SyRI, ¿a quién sanciono?' Garantías frente al uso de inteligencia artificial y decisiones automatizadas en el sector pú-

- blico y la sentencia holandesa de febrero de 2020, en *La Ley Privacidad* , *Wolters Kluwer* n° 4, mayo. Acceso en Academia
- COTINO HUESO, L. (2020 b): "El alcance e interactuación del régimen jurídico de los datos personales y big data relacionado con salud y la investigación biomédica", Revista de derecho y genoma humano: genética, biotecnología y medicina avanzada, nº 52, pp. 57-96, nº52 enero-junio 2020, acceso
- COTINO HUESO, L. (2021 a): "Hacia la transparencia 4.0: el uso de la inteligencia artificial y big data para la lucha contra el fraude y la corrupción y las (muchas) exigencias constitucionales", en Carles RAMIÓ (coord.), Repensando la administración digital y la innovación pública, Instituto Nacional de Administración Pública (INAP), Madrid, 2021. https://links.uv.es/FUW2pz6
- COTINO HUESO, L. (2021 b): "El nuevo reglamento de Administración electrónica, que no innova en tiempos de transformación digital", *Revista Catalana de Derecho Público* nº 63, 2021, pp. 118-136, DOI: http://dx.doi.org/10.2436/rcdp.i63.2021.3672
- COTINO HUESO, L. (2022 a): "Nuevo paradigma en la garantías de los derechos fundamentales y una nueva protección de datos frente al impacto social y colectivo de la inteligencia artificial", en COTINO HUESO, Lorenzo (editor), *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*, Thompson-Reuters Aranzadi, FIADI (Federación Iberoamericana de Asociaciones de Derecho e Informática), Cizur.
- COTINO HUESO, L. (2022 b): "Derechos ante la administración digital y la inteligencia artificial (XVIII Y XV)", en COTINO HUESO, Lorenzo (editor), *La Carta de Derechos Digitales*, Tirant Lo Blanch, Valencia, 2022, pp. 251-284.
- COTINO HUESO, L. (editor) (2022), *La Carta de Derechos Digitales*, Tirant Lo Blanch, Valencia.
- COTINO HUESO, L. y otros (2021): "Un análisis crítico constructivo de la Propuesta de Reglamento de la Unión Europea por el que se establecen normas armonizadas sobre la Inteligencia Artificial (Artificial Inteligence Act)", *Diario La Ley*, 2 de julio, Wolters Kluwer. https://links.uv.es/2FK3xc4
- CRAWFORD, K. Y SCHULTZ, J. (2014): "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms". *Boston College Law Review, Vol. 55, No. 93* https://ssrn.com/abstract=2325784
- DE MIGUEL BERIAIN, I. (2018). "Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling, *Law, Probability and Risk, Volume 17, Issue 1, 1* March 2018, pp. 45–53, https://doi.org/10.1093/lpr/mgy001

DE LA SIERRA MORÓN, S. (2020): "Inteligencia artificial y justicia administrativa: una aproximación desde la teoría del control de la Administración Pública", *Revista General de Derecho Administrativo*, nº. 53, 2020, https://laadministracionaldia.inap.es/noticia.asp?id=1510413

- DELGADO GOBOY y otras (2003): Guía de aplicación práctica, para la elaboración de informes de impacto de género de las disposiciones normativas que elabore el gobierno, de acuerdo a la ley 30/2003, 2003, Instituto De La Mujer (Ministerio de Trabajo Y Asuntos Sociales), https://www.inmujeres.gob.es/publicacioneselectronicas/documentacion/Documentos/DE0259.pdf
- EUROPEAN INSTITUTE OF GENDER EQUALITY (EIGE), Artificial intelligence, platform work and gender equality, 2021, https://eige.europa.eu/publications/artificial-intelligence-platform-work-and-gender-equality-report
- EUROPEAN LAW INSTITUTE (2022): Model Rules on Impact Assessment of Algorithmic Decision-Making Systems Used by Public Administration, European Law Institute-University of Viena, https://www.europeanlaw-institute.eu/fileadmin/user\_upload/p\_eli/Publications/ELI\_Model\_Rules\_on\_Impact\_Assessment\_of\_ADMSs\_Used\_by\_Public\_Administration.pdf
- EXECUTIVE OFFICE OF THE PRESIDENT (2016): Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights. May, 2016. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016\_0504\_data\_discrimination.pdf
- FEATHERS, T. (2021): "Major Universities Are Using Race as a "High Impact Predictor" of Student Success: Students, professors, and education experts worry that that's pushing Black students in particular out of math and science", en *The Markup*. Marzo https://themarkup.org/machinelearning/2021/03/02/major-universities-are-using-race-as-a-high-impact-predictor-of-student-success
- FRA (Agencia de la Unión Europea para los Derechos Fundamentales) (2018): UE. (2018). #BigData: Discrimination in data-supported decision making, European Union Agency for Fundamental Rights, Viena, https://fra.europa.eu/en/publication/2018/big-data-discrimination
- FRA, (2019): Data quality and artificial intelligence- mitigating bias and error to protect fundamental rights, Luxembourg, Publications Office, junio, https://fra.europa.eu/en/publication/2019/artificial-intelligence-data-quality
- FRA (2020): Artificial Intelligence and Fundamental Rights, Agencia de los Derechos Fundamentales de la Unión Europea, 2020, https://fra.europa.eu/sites/default/files/fra\_uploads/fra-2020-artificial-intelligence\_en.pdf

- FRA (2021): Getting the future right artificial intelligence and fundamental rights, FRA, https://fra.europa.eu/en/publication/2020/artificial-intelligence-and-fundamental-rights
- FRA- CONSEJO DE EUROPA (2018): Manual de legislación europea contra la discriminación, revisión 2018 340 pp., Oficina de Publicaciones de la Unión Europea, https://www.echr.coe.int/documents/handbook\_non\_discri\_law\_spa.pdf
- FREEMAN ENGSTROM, D. y otros (2020): "Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies. Report submitted to the Administrative conference of the United States", NYU School of Law, Public Law Research Paper No. 20-54, 122 págs., http://dx.doi.org/10.2139/ssrn.3551505
- FTC, Comisión Federal del Comercio (2016). Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues (6 enero) . https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report
- GAMERO CASADO, E. (2021): "Necesidad de motivación e invalidez de los actos administrativos sustentados en inteligencia artificial o en algoritmos", *Almacén de Derecho*, 4 febrero, https://almacendederecho.org/necesidad-de-motivacion-e-invalidez-de-los-actos-administrativos-sustentados-en-inteligencia-artificial-o-en-algoritmos
- GAMERO CASADO, E. y FERNÁNDEZ RAMOS, S. (2016): Manual básico de Derecho Administrativo, (13ª ed.), Tecnos, 2016
- GIMÉNEZ GLUCK, D. (2004): *Juicio de igualdad y Tribunal Constitucional*, J.M. Bosch Editor, Barcelona.
- GIMÉNEZ GLUCK, D. (1999): Una manifestación polémica del principio de igualdad: acciones positivas moderadas y medidas de discriminación inversa, Tirant lo Blanch, Valencia.
- GONZÁLEZ-TREVIJANO, P. (2020): Los principios de igualdad y no discriminación, una perspectiva de Derecho Comparado. España, EPRS Servicio de Estudios del Parlamento Europeo, PE 659.297, Octubre, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/659297/EPRS STU(2020)659297 ES.pdf
- GOODMAN, B. W. (2016): "A step towards accountable algorithms?: Algorithmic discrimination and the European Union general data protection", 29th Conference on Neural Information Processing Systems, http://www.mlandthelaw.org/
- GRUPO DEL ARTÍCULO 29 (2017): Directrices sobre la evaluación del impacto de la protección de datos (DPIA) y determinar si el procesamiento es "probable que genere un alto riesgo" a los efectos del Reglamento 2016/679, WP248 rev.01, Revisadas por última vez y adoptadas

el 4 de octubre de 2017, https://www.aepd.es/sites/default/files/2019-09/wp248rev01-es.pdf

- GRUPO DEL ARTÍCULO 29 (2018): Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679. 3 de octubre de 2017 y revisadas el 6 de febrero de 2018.
- GUTIÉRREZ DAVID, M.E., "Acceso al código fuente y a los algoritmos de las Administraciones inteligentes. Lecciones a partir de experiencias comparadas", en COTINO HUESO L. y CASTELLANOS CLARAMUNT J. (coords.) (2022), *ob. cit.*, pp.135-167.
- HERNÁNDEZ PEÑA, J. C. (2022): El marco jurídico de la inteligencia artificial. Principios, procedimientos y estructuras de gobernanza, Aranzadi, Cízur.
- HERMSTRÜWER, Y. (2020): "Artificial Intelligence and Administrative Decisions Under Uncertainty", en WISCHMEYER, T y RADEMACHER. T. (ed.), (2020), *ob. cit.* pp. 199-223.
- HLEG (Comisión Europea–Grupo Independiente de Expertos de Alto nivel sobre Inteligencia Artificial), (2019): *Directrices éticas para una IA fiable*, 2019, https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1
- HUERGO LORA, A. J. (2020): "Una aproximación a los algoritmos desde el derecho administrativo", en HUERGO LORA, A. J. (dir.), DÍAZ GONZÁLEZ, G. M. (coord.) *La regulación de los algoritmos*, Aranzadi Thomson Reuters, Cízur, 2020.
- HUERGO LORA, A. J. (2021): "Administraciones Públicas e inteligencia artificial: ¿más o menos discrecionalidad?", *El Cronista del Estado Social y Democrático de Derecho*, nº 96-97, 2021, pp. 78-95, sigo por https://laadministracionaldia.inap.es/noticia.asp?id=1512494
- ISO (2006): Statistics Vocabulary and symbols Part 1: General statistical terms and terms used in probability, ISO, Tech. Rep. ISO 3534-1:2006, https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/04/01/40145.html
- JAUME PALASÍ, L. (2020): "Cómo la inteligencia artificial está impactando en las sociedades", Cerrillo I Martínez, A. Y Peguera Poch, M. (coords.), (2020): Retos jurídicos de la inteligencia artificial, Aranzadi, Cizur, pp. 27-39.
- JIMÉNEZ LÓPEZ, Jesús (2022): "Oscuridad algoritmica en el sector público", en VESTRI, G. (2022) *ob. cit.* pp. 41-66.
- JOH, E. E. (2017): "The Undue Influence of Surveillance Technology Companies on Policing" (27 febrero). N.Y.U. L. Review Online. https://ssrn.com/abstract=2924620
- JOHNSON, C. (2022 a): Flaws plague a tool meant to help low-risk federal prisoners win early release. NPR. 26 enero, 2022 https://www.npr.

- org/2022/01/26/1075509175/flaws-plague-a-tool-meant-to-help-low-risk-federal-prisoners-win-early-release;
- JOHNSON, C. (2022 b): Justice Department works to curb racial bias in deciding who's released from prison. NPR. 19 de abril, https://www.npr. org/2022/04/19/1093538706/justice-department-works-to-curb-racialbias-in-deciding- whos-released-from-pris
- JOINT RESEARCH CENTRE, COMISIÓN EUROPEA (2022): AI Watch, road to the adoption of artificial intelligence by the public sector: a handbook for policymakers, public administrations and relevant stakeholders, Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2760/288757
- JOINT RESEARCH CENTRE, ESTÉVEZ ALMENZAR, M. y otros (2022): Glossary of human-centric artificial intelligence, Joint Research Centre, Unión Europea, Luxemburgo, https://publications.jrc.ec.europa.eu/repository/handle/JRC129614 DOI:10.2760/860665
- JOINT RESEARCH CENTRE, MISURACA, G. (2020), AI Watch. Artificial Intelligence in public services. Overview of the use and impact of AI in public services in the EU, Joint Research Centre, Unión Europea. https://publications.jrc.ec.europa.eu/repository/bitstream/JRC120399/jrc120399\_misuraca-ai-watch\_public-services\_30062020\_def.pdf
- JOINT RESEARCH CENTRE, TANGI, L. y otros (2022): AI Watch European landscape on the use of Artificial Intelligence by the Public Sector, JRC Science For Policy Report, Publications Office of the European Union, Luxembourg, 2022, doi:10.2760/39336, https://publications.jrc.ec.europa.eu/repository/handle/JRC129301
- JORDANO FRAGA, J. (1996): "Nulidad de los actos administrativos y derechos fundamentales, Marcial Pons, Madrid.
- JORDANO FRAGA, J. (2011): "La nulidad de los actos que lesionen derechos fundamentales: evolución jurisprudencial y desarrollo de la causa de nulidad, nucleo de las garantías en el Estado de Derechos Fundamentales", en LÓPEZ MENUDO F. (coord.), Derechos y garantías del ciudadano: estudios en homenaje al profesor Alfonso Pérez Moreno, 2011, pp. 715-743.
- KAMMOURIEH, L y otros, "Group privacy in the age of big data", en TAYLOR. L.; VAN DER SLOOT, B.; FLORIDI, L. (eds.) (2017): *Group Privacy: New Challenges of Data Technologies*, Springer, Capítulo 3, pp. 48-83, https://www.stiftung-nv.de/sites/default/files/group-privacy-2017-authors-draft-manuscript.pdf
- LAMMERANT, H. y otros (2019): Big data decision-making processes and shortcuts of discrimination, NTM/NJCM-bull. 2018/1
- LIKADI (2018): Guía para la evaluación de impacto de género de la normativa guía para la evaluación de impacto de género de la norma-

tiva, Diputación Foral de Álava, 2018, https://web.araba.eus/documents/105044/1393927/Gu%C3%ADapara+la+evaluaci%C3%B3n+de+impacto+de+g%C3%A9nero+%C3%81lava+castellano+navegable.pdf/74905ae0-fc50-7b15-a5c3-2ab4b2f16fa3?t=1557493191803

- LOZA CORERA, M. (2022): "Explotación y regulación del uso del big data e inteligencia artificial para los servicios públicos y la ciudad inteligente", en COTINO HUESO, L. y TODOLÍ SIGNES, A. (coord.), *ob. cit.* pp. 125-148.
- MANTELERO, A. (2017): "From group privacy to collective privacy: towards a new dimension of privacy and data protection in the big data era", en TAYLOR. L.; VAN DER SLOOT, B.; FLORIDI, L. (eds.) ob. cit. Capítulo 8.
- MANTELERO, A. (2022): Beyond Data. Human Rights, Ethical and Social Impact Assessment, Springer, Information Technology and Law Series IT&LAW 36, https://link.springer.com/book/10.1007/978-94-6265-531-7
- MARTÍN DELGADO, I. (2009): "Naturaleza, concepto y régimen jurídico de la actuación administrativa automatizada", *Revista de administración pública*, nº 180, pp. 353-386.
- MARTÍNEZ GARAY, L. y MONTES SUAY, F. (2018): "El uso de valoraciones del riesgo de violencia en Derecho Penal:: algunas cautelas necesarias", *Indret: Revista para el Análisis del Derecho*, nº. 2,.
- MARTÍNEZ MARTÍNEZ, R. (2019): "Inteligencia artificial desde el diseño. Retos y estrategias para el cumplimiento normativo", *Revista catalana de dret públic*, nº 58, pp. 64-81.
- MELLADO RUIZ, L. (2017), "Supuestos de nulidad de pleno derecho de los actos administrativos", en GAMERO CASADO, E. (dir.), *Tratado de Procedimiento Administrativo Común y Régimen Jurídico Básico del sector público* Vol. 2, Tirant lo Blanch, Valencia, pp. 1929-1983.
- MINISTERIE VAN BINNENLANDSE ZAKEN; VAN DER SLOOT, B. y otros (2021): *Non-discriminatie by design*, 69 pp., encargo para (Ministerio del Interior), Tweede Kamer (Cámara de representantes), https://www.tweedekamer.nl/downloads/document?id=2021D22772
- MINISTERIE VAN JUSTITIE EN VEILIGHEID (Ministerio de Justicia y Seguridad), Rijksoverheid (Gobierno central) (2021), Richtlijnen voor het toepassen van algoritmen door overheden en publieksvoorlichting over data-analyses (Pautas para la aplicación de algoritmos por parte de los gobiernos y educación pública sobre análisis de datos), Directiva (Richtlijn), de 08-03-2021, https://acortar.link/6C226N
- MINISTRY OF THE INTERIOR AND KINGDOM (2022): *Impact Assessment. Fundamental rights and algorithms (IAMA)*, Países Bajos, marzo. https://www.government.nl/binaries/government/documenten/

- reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms/Fundamental+Rights+and+Algorithms+Impact+Assessment.pdf
- MIRÓ LLINARES, F. (2022): "Inteligencia artificial, delito y control penal: nuevas reflexiones y algunas predicciones sobre su impacto en el derecho y la justicia penal", en *El Cronista del Estado Social y Democrático de Derecho*, nº. 100, 2022, pp. 174-183.
- MORAL SORIANO, L. (2022): "Inteligencia artificial y filosofía del derecho decisiones automatizadas, derecho administrativo y argumentación jurídica", en LLANO ALONSO F. H. (dir.), *Inteligencia artificial y filosofía del derecho*, Laborun, Madrid, 2022, pp. 475-498. https://idus.us.es/bitstream/handle/11441/137250/Inteligencia%20artificial\_Llano%20Alonso.pdf?sequence=1&isAllowed=y
- NATIONAL INSTITUTE OF JUSTICE (2021): Review and Revalidation of the First Step Act Risk Assessment Tool. National Institute of Justice NCJ 303859. https://www.ojp.gov/pdffiles1/nij/303859.pdf
- NAVAS NAVARRO, S. (2017): "Derecho e inteligencia artificial desde el diseño. Aproximaciones", NAVAS NAVARRO, S. (coord..). *Inteligencia artificial: tecnología, derecho*, Tirant lo Blanch, Valencia, pp. 23-72.
- NICOLAS, T.(2014) "Big Data Proxies and Health Privacy Exceptionalism", en *Health Matrix n*° 24, pp. 65-108.
- NIST (2022): AI Risk Management Framework: Initial Draft, 17 de marzo, https://www.nist.gov/document/ai-risk-management-framework-initial-draft
- NIST; SCHWARTZ, R. y otros (2022): Towards a Standard for Identifying and Managing Bias in Artificial Intelligence, Special Publication 1270 (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, https://doi.org/10.6028/NIST.SP.1270
- NOORDT, C., y MISURACA, G. (2022), "Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union", *Government Information Quarterly*, 101714, https://doi.org/10.1016/j.giq.2022.101714
- O'NEIL, C. (2017): Armas de Destrucción Matemática. Como el Big Data aumenta la desigualdad y amenaza la democracia. Capitán Swing, Madrid.
- OBERMEYER, Z. y otros (2019), Dissecting racial bias in an algorithm used to manage the health of populations. Science. Vol. 366, No. 6464. Oct. 25, 2019 https://www.science.org/doi/10.1126/science.aax2342
- PALMA ORTIGOSA, A. (2021): Régimen jurídico de la toma de decisiones automatizadas y el uso de sistemas de inteligencia artificial en el marco del derecho a la protección de datos personales, Tesis doctoral Universidad de Valencia, 2021

PARLAMENTO EUROPEO (2017): Resolución de 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley (2016/2225(INI)). Acceso en http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0076+0+DOC+XML+V0//ES

- POLO ROCA, A. (2021): "Datos, datos, datos: el dato personal, el dato no personal, el dato personal compuesto, la anonimización, la pertenencia del dato y otras cuestiones sobre datos". *Estudios de Deusto: revista de la Universidad de Deusto*, Vol. 69, nº. 1, pp. 165-194, DOI: 10.18543/ed-69(1)-2021
- RED DAIA (2020 a): Aportación a consulta de derechos digitales de julio, 2020 https://acortar.link/q1wl3i
- RED DAIA (2020 b): Aportación DAIA a consulta pública Carta de derechos digitales, diciembre 2020 https://bit.ly/3paF0H0
- REY MARTÍNEZ, F. (2017): Igualdad y prohibición de discriminación: de 1978 a 2018, Revista de derecho político, Nº 100, 2017, pp. 125-171.
- REY MARTÍNEZ, F. (2020): Derecho antidiscriminatorio, Aranzadi Thomson Reuters, Cízur.
- ROEHL, U. (2022): "Understanding Automated Decision-Making in the Public Sector: A Classification of Automated, Administrative Decision-Making", en Juell-Skielse, G. y otros (eds.), Service Automation in the Public Sector. Progress in IS. Springer, Cham, pp. 35-63, DOI: 10.1007/978-3-030-92644-1\_3,
- RUBIO LLORENTE, F. (1993): "Igualdad en la jurisprudencia del Tribunal Constitucional. Introducción", en *La Forma del Poder (Estudios sobre la Constitución)*, CEC, Madrid, pp. 637-684;
- SANGEN, M. (2019), Research on fair algorithms for policy, 6-8-2019, https://www.cbs.nl/nl-nl/corporate/2019/32/research-on-fair-algorithms-for-policy.
- SIMMONS, R. (2016): "Quantifying Criminal Procedure: How to Unlock the Potential of Big Data in Our Criminal Justice System", *Mich. St. L. Rev.* 947 (Ohio State Public Law Working Paper No. 362. pp. 977 y ss. http://dx.doi.org/10.2139/ssrn.2816006
- SIMÓ, P. y MAGRO, V. (2021): Justicia cautelar e inteligencia artificial: la alternativa a los atávicos heurísticos judiciales, J.M. Bosch Editor, Barcelona, 2021.
- SIMÓ, P. (2022): La prisión algorítmica, Tirant lo Blanch, Valencia.
- SORIANO ARNANZ, A. (2021 a): "Decisiones automatizadas y discriminación: aproximación y propuestas generales", *Revista General de Derecho Administrativo* (Iustel, enero 2021) nº 56, acceso, https://laadministracionaldia.inap.es/noticia.asp?id=1511706

- SORIANO ARNANZ, A. (2021 b): "Decisiones automatizadas: problemas y soluciones jurídicas. Más allá de la protección de datos", *Revista de Derecho Público: Teoría y Método*, Vol. 3, pp. 85-127, DOI:10.37417/RPD/vol 1 2021 535
- SORIANO ARNANZ, A. (2022 a): "El uso de sistemas automatizados por los poderes públicos. El caso del sistema empleado por el servicio público de empleo austriaco", en COTINO HUESO, L. y TODOLÍ SIGNES, A. (coord.), Explotación y regulación del uso del big data e inteligencia artificial para los servicios públicos y la ciudad inteligente, Tirant lo Blanch, Valencia, pp. 192-216.
- SORIANO ARNANZ, A. (2022 b): "Discriminación algorítmica: garantías y protección jurídica", en COTINO HUESO, L. (coord.), *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*, Aranzadi, Cizur Menor, 2022, págs. 139-169.
- TISCHBIREK, A. (2020): "Artificial Intelligence and Discrimination: Discriminating Against Discriminatory Systems", en WISCHMEYER, T y RA-DEMACHER. T. (ed.), (2020), *ob. cit.* pp. 103-121.
- U.S. DEPARTMENT OF JUSTICE (2022): Algorithms, Artificial Intelligence, and Disability Discrimination in Hiring. Mayo, https://beta.ada.gov/resources/ai-guidance/
- U.S. EQUAL EMPLOYMENT OPPORTUNITY COMMISSION (2022): The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees. EEOC-NVTA-2022-2, https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence
- U.S. TRANSPORTATION SECURITY ADMINISTRATION (2022): *Transgender/ Non Binary / Gender Nonconforming Passengers*. TSA, https://www.tsa.gov/transgender-passengers
- UNESCO (2020): Artificial Intelligence and Gender Equality. Key findings of UNESCO's Global Dialogue, 2020, https://unesdoc.unesco.org/ark:/48223/pf0000374174
- VALERO TORRIJOS, J. (2007): El régimen jurídico de la e-Administración. El uso de los medios informáticos y telemáticos en el procedimiento administrativo común, 2.ª ed., Comares, Granada.
- VALERO TORRIJOS, J. (2013): Derecho, innovación y administración electrónica, Global Law = Derecho Global, Sevilla.
- VALERO TORRIJOS, J. (2015): "Ciudades inteligentes y datos abiertos: implicaciones jurídicas para la proteccion de los datos de carácter personal", *Istituzioni del federalismo: rivista di studi giuridici e politici*, n°. 4, 2015 ("Smart Cities" e amministrazioni intelligenti), pp. 1025-1047.

VALLE ESCOLANO, R., "Transparencia en la inteligencia artificial y en el uso de algoritmos: una visión de género", COTINO HUESO L. y CASTE-LLANOS CLARAMUNT J. (coords.) (2022), *ob. cit.*, pp. 85-110.

- VAN DER SLOOT, B. (2021): "Regulating non-personal data in the age of Big Data", en TZANOU, M. (Ed.) (2021): Health data privacy under the GDPR: Big Data challenges and regulatory responses, Routledge pp. 85-105; p. 16 de la versión SSRN.
- VELASCO RICO, C. I. (2022): "Tecnologías disruptivas en la Administración Pública. Inteligencia artificial y Blockchain", en CERRILLO I MARTÍNEZ, A. (coord.) *La Administración Digital*, Dykinson, Madrid, pp. 228-256.
- VESTRI, G. (2021): "Sistemas de inteligencia artificial en la contratación pública: entre códigos fuente y datos abiertos", *Actualidad administrativa*, nº 12.
- VESTRI, G. (2022 a): "Denegación vs. Derecho de acceso al código fuente en los sistemas algorítmicos. Una perspectiva jurídico-administrativa", COTINO HUESO L. y CASTELLANOS CLARAMUNT J. (coords.) (2022), ob. cit., pp. 111-134.
- VESTRI, G. (2022 b): Disrupción tecnológica en la administración pública retos, Aranzadi, Cízur.
- WACHTER, S. (2022): "The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law", 15 de febrero, *Tulane Law Review*, http://dx.doi.org/10.2139/ssrn.4099100
- WACHTER, S. y MITTELSTADT, B. D. (2019): "A right to reasonable inferences: re-thinking data protection law in the age of big data and AI", *Columbia Business Law Review*, vol. 2019, No. 2, pp. 494-620, versión https://ssrn.com/abstract=3248829
- WACHTER, S., y otros, "Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law", West Virginia Law Review, Vol. 123, núm. 735, 2021, https://x.gd/hKKKqS
- WHITE HOUSE (2022): The Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People, White House Office of Science and Technology Policy, Whasington DC., octubre, https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf
- WISCHMEYER, T y RADEMACHER. T. (ed.), (2020): Regulating Artificial Intelligence, Springer, 2020, https://doi.org/10.1007/978-3-030-32361-5
- XENIDIS, R. y SENDEN, L. (2020): "EU Non-discrimination law in the era of artificial intelligence: mapping the challenges of algorithmic discrimination", en BERNITZ, U. y otros, (eds.), *General Principles of EU Law and the EU Digital Order*, Alphen aan den Rijn, Kluwer Law International, 2020.

ZLOTNIK, A. (2019): "Inteligencia Artificial en las Administraciones Públicas: definiciones, evaluación de viabilidad de proyectos y áreas de aplicación", *Boletic*, núm. 84, pp. 27-28, http://apiscam.blogspot.com/2019/11/boletic-revista-de-astic-monografico-de.html