PROBLEMAS ÉTICO-JURÍDICOS DE LAS DECISIONES ALGORÍTMICAS Y EL BIG DATA*

Juan C. Hernández
UNED Tudela
jchernandez@tudela.uned.es

Joslay Polanco Medina
Univ. Internacional de La Rioja
joslay. polanco@unir.net

Recepción: 1 de diciembre de 2019; Aceptación: 7 de enero de 2020.
Citación APA: Hernández, Juan C.; Polanco Medina, Joslay (2020). «Problemas ético-jurídicos de las decisiones algorítmicas y el Big data». *Revista de Humanidades Cuadernos del Marqués de San Adrián, n.º 12*, UNED Tudela, pp. 75-108.

Resumen:
En este trabajo se analizan algunos de los problemas más relevantes respecto a las decisiones algorítmicas y el Big Data. En primer lugar, se realiza una aproximación respecto a la inteligencia artificial, así como un análisis de la regulación que a la fecha se ha desarrollado en el marco del Consejo de Europa. Posteriormente, se abordan los principales problemas del perfilado y las decisiones algorítmicas en la regulación europea de protección de datos, las peculiaridades del perfilado ideológico, y finalmente, la inacabada regulación de las discriminaciones algorítmicas.

Palabras clave: Big Data, decisiones algorítmicas, perfilado, perfilado ideológico, discriminaciones algorítmicas.

Abstract:
In this paper we analyse some of the most relevant problems regarding algorithmic decisions and Big Data. Firstly, an approximation is made with respect to artificial intelligence, as well as an analysis of the regulation that has been developed to date within the framework of the Council of Europe. Subsequently, the main problems of profiling and algorithmic decisions in European data protection regulation are studied, the peculiarities of ideological profiling, and finally, the incomplete regulation of algorithmic discriminations.

Keywords: Big Data, algorithmic decisions, profiling, ideological profiling, algorithmic discriminations.

* La investigación que ha dado lugar a estos resultados ha sido impulsada por la Obra Social «La Caixa», y la Fundación Bancaria Caja Navarra, en colaboración con el Centro Asociado a la UNED de Tudela.

Résumé :
Dans cet article, nous analysons certains des problèmes les plus pertinents concernant les décisions algorithmiques et les Big Data. Tout d'abord, une approximation est faite en ce qui concerne l'intelligence artificielle, ainsi qu'une analyse de la réglementation qui a été élaborée jusqu'à présent dans le cadre du Conseil de l'Europe. Par la suite, les principaux problèmes du profilage et des décisions algorithmiques dans la réglementation européenne sur la protection des données, les particularités du profilage idéologique et, enfin, la réglementation inachevée des discriminations algorithmiques sont abordés.

Mots-clés : Big Data, décisions algorithmiques, profilage, profilage idéologique, discriminations algorithmiques.

I. Introducción

En junio de 2014 se publicó en los *Proceeding of the National Academy of Science* de Estados Unidos, un artículo de tres investigadores vinculados a la Universidad de Cornell y al equipo de investigación de Data Science de Facebook. El artículo, titulado *Experimental evidence of massive-scale emotional contagion through social networks*[1], fue el más citado del año 2014, así como objeto de un elevado número de críticas desde distintos flancos, y documenta un controvertido experimento social realizado con fines de determinar el alcance del efecto contagio en las redes sociales manipulando, para ello, algunos algoritmos de Facebook[2].

Las críticas se centraron principalmente en el desconocimiento de los usuarios acerca de su participación en un estudio científico con impacto sobre la salud (se manipulaban emociones), y la justificación dada por los investigadores de que el consentimiento informado no era necesario por estar incluido de manera expresa en los términos de aceptación del servicio que los usuarios suscriben al momento de crear una cuenta en la red social. Sin embargo, el problema colateral que desveló el estudio, y que representa una de las preocupaciones de este proyecto de investigación, es que las noticias y en definitiva toda la información que reciben los usuarios vía algoritmo de Facebook, puede ser objeto de fácil manipulación sin que los usuarios seamos conscientes de ello.

Por otra parte, en marzo de 2018, el *New York Times* y *The Guardian*, develaron el conocido caso *Facebook-Cambridge Analytica*, que develó la utilización no autorizada de datos de más de 50 millones de usuarios de esa red social, con el objetivo de perfilarlos políticamente y enviarles información que algunos han calificado como *Fake News*. Más allá de la verdadera efectividad de esta estrategia electoral, el escandalo puso en evidencia la vulnerabilidad respecto a los datos que ingenuamente los usuarios depositan en la red social, y la posibilidad cierta de extraer información latente con miras a desarrollar perfiles que afectan información habitualmente calificada como sensible por la normativa de protección de datos.

Ambos casos, así como otros que obviamos, han puesto en evidencia varios aspectos de relevancia social, que merecen ser estudiados desde una perspectiva ético-jurídica respecto a la mediatización de algoritmos en la vida social.

El elemento que retícula esta nueva problemática, y que está a su vez en el centro de este proyecto de investigación, es el incremento exponencial a decisiones algorítmicas que pueden tener un impacto real sobre las personas y que van desde el posible acceso o denegación de servicios que consideramos indispensables para desarrollar una vida digna (ej. la calificación crediticia para acceder a un préstamo hipotecario) hasta la afectación en el ejercicio de derechos fundamentales, basado -por citar sólo un par de casos- en discriminaciones inadvertidas por los desarrolladores de los algoritmos, o la creación de perfiles que permita a grupos políticos manipular la opinión de votantes. La cara oculta de esta problemática la representa el negocio en auge de la explotación de datos personales (es por todos conocida la frase atribuida al matemático Clive Humby: *the data is the new oil*), así como la escasamente debatida subjetividad de los algoritmos.

En definitiva, el auge del Big Data, el *Machine Learning* y la Inteligencia Artificial, han potenciado tanto las ventajas como los posibles riesgos para la sociedad de las decisiones algorítmicas. Tanto Gobiernos como diversas instituciones pueden encontrar en ellos una herramienta valiosa para adoptar decisiones en un mundo cada vez más complejo por la acumulación masiva de datos, que podrían aprovecharse para mejorar la actuación pública y desarrollar políticas más efectivas y eficientes en

1 GUILLORY, Guillory, J. E. *et al.* (2014), «Experimental evidence of massive scale emotional contagion through social networks», *Proceedings of the National Academy of Sciences*, 111(29), pp.10779–10779.

2 De manera más concreta, el experimento buscaba determinar si los estados emocionales pueden ser objeto de transferencia a escala masiva a través de las redes sociales. Los investigadores buscaban comprobar si el efecto contagio, positivo y negativo, ampliamente aceptado y comprobado en estudios sociales de redes, se replicaría en las redes sociales utilizando Facebook. Para ello manipularon la exposición de más de 600 000 usuarios de la red social a través del *News Feed*.

ámbitos urgentes y de profundo impacto social[3]. Sin embargo, tal como señalaron sendos informes ejecutivos preparados por la Oficina de Presidencia de Estados Unidos[4], las decisiones algorítmicas también pueden encubrir situaciones de discriminación relacionadas, e. g., con oportunidades de acceso a servicios o respecto al ejercicio de derechos fundamentales vinculados al empleo, vivienda, educación o incluso principios tan fundamentales como la presunción de inocencia.

Por tanto, el punto sobre el cuál debe pivotar todo el análisis es que una vez que hemos puesto el enfoque de estas técnicas sobre las personas, y no únicamente sobre los productos, es necesario robustecer el marco de análisis ético-jurídico, incluyendo, como no puede ser de otra manera, los derechos humanos y fundamentales como referencia condicionante de su marco regulatorio.

En el mundo jurídico hispanohablante, no obstante, la atención a esta problemática es prácticamente inexistente, pese a que adquiere nueva relevancia ante la entrada en vigor del Reglamento General de Protección de Datos (RGPD) en mayo de este año, así como la reciente publicación de la Ley Orgánica de Protección de Datos y Garantía de los Derechos Digitales, que debemos escrutar a efectos de determinar si dan respuesta a estos problemas o requieren mayor reflexión y propuestas de mejoras.

Los objetivos de este trabajo no dejan de ser modestos. Se dirigen, por una parte, a identificar algunos problemas de calado, sin agotar las aristas posibles. Por otra, presenta la regulación y soluciones jurídicas vigentes, dejando constancia de algunas limitaciones.

De manera concreta, se pretende abordar el problema de las decisiones automáticas o algorítmicas, en segundo término, hacer referencia al problema del perfilado ideológico, y finalmente acercarnos a la problemática ético-jurídica de las discriminaciones adoptadas por mediación de algoritmos. Todo esto se contextualizará

No obstante, como el Derecho en general, y el Derecho de la tecnología en particular está abierto a factores extrajurídicos, resulta necesario presentar una breve explicación acerca de lo que significa inteligencia artificial, específicamente Machine Learning, y Big Data.

3 Un análisis general sobre las posibilidades de estas tecnologías, y su problemática en la administración pública, en COGLIANESE, C. & LEHR, D. (2017), *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*.

4 PODESTA, J. *et al.* (2014), «Big Data: Seizing Opportunities», *Executive Office of the President of USA*, p. 51; MUNOZ, C., *et. al*, (2016), «Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights». *Executive Office of the President of USA*.

En este sentido, como señalara Hoffman-Riem[5], sin comprender la realidad a las que las normas se refieren dificulta entender el panorama normativo y la respuesta del legislador a la problemática que pretende solucionar.

II. Inteligencia artificial y Big Data: Una aproximación para juristas e investigadores sociales

La inteligencia artificial, que puede tener como objetivo final la creación de entes autónomos capaces de aprender y tomar decisiones complejas, en ambientes cambiantes y se encuentren habilitados para transmitir conocimientos, decisiones y pensamientos emulando a los humanos, dista de ser realidad.

A partir del desarrollo tecnológico actual, es razonable sostener que esta inteligencia artificial «dura» no es propiamente una quimera, pero si un hito que aún escapa al conocimiento humano. Por tanto, conviene tener presente que cuando se hace referencia a aplicaciones relacionadas con inteligencia artificial, habitualmente nos estamos refiriendo a *Machine Learning*.

El concepto de *Machine Learning* fue acuñado por Arthur Samuel en 1959. Lo hizo en su pionero estudio acerca de un programa de ordenador que aprendía a jugar al ajedrez y terminaría venciéndole[6]. Desde ese momento, este campo de las ciencias computacionales se asocia con la capacidad de aprendizaje automático de los programas, sin que sean minuciosamente programados para ello. Aunque esta definición inicial no sea la más acertada, apuntala una de sus características distintivas: la capacidad de aprender de la experiencia, rememorando así el aprendizaje del ser humano[7].

El *machine learning*, o aprendizaje automático, podemos entenderlo como un proceso que permite a los programas extraer patrones o realizar

5 HOFFMAN-RIEM, W., (2004), «Methoden einer anwendungsorientierten Verwaltungsrechtswissenschaft», en SCHMIDT-ASSMANN, E. y HOFFMAN-RIEM, W. (ed.), *Methoden der* Verwaltungsrechtswissenschaft, Nomos, Baden, p. 36.

6 SAMUEL, Arthur L. (1959), «Some Studies in Machine Learning Using the Game of Checkers», en *IBM Journal*, Vol. 3, Núm. 3.

7 El Machine Learning está estrechamente relacionado, y se suele identificar como una versión débil de la inteligencia artificial. Una explicación más detallada en NEAPOLITAN, R.E y JIANG, X. (2018), *Artificial Intelligence. With an Introduction to Machine Learning*, segunda edición, CRC Press, Boca Ratón, pp. 2-3. También es canónico el trabajo de RUSSELL, S. Y NORVIG, P. (2016), *Artificial Intelligence. A Modern Approach*, tercera edición, Pearson, Essex.

asociaciones a partir de conjuntos de datos[8], que resultan inescrutables para el ser humano. En otras palabras, se recurre a algoritmos que adquieren experiencia o aprenden, con escasa o muy limitada intervención humana, a partir de los ejemplos que le son suministrados. Esa nueva experticia la aplican a nueva información, a fin de resolver un problema dado o hacer predicciones, ya sea determinar si un correo es *spam*, un cliente pagará una hipoteca en atención a su perfil o el precio más acorde de un piso a partir de sus características.

El *machine learning* se encuentra en la intersección de muchas disciplinas, como la neurociencia, la psicología, la cibernética, la lingüística o la economía, pero se cimienta fundamentalmente en las ciencias computacionales y estadísticas[9]. Aunque su funcionamiento es complejo y envuelve un cierto número de fases, cómo todo proyecto de analítica avanzada, podemos resumirlo en que usualmente se utilizan datos históricos que sirven de ejemplo para entrenar los algoritmos, y a partir de los cuales extraen los resultados del modelo, además —en buena medida— de los parámetros. No obstante, el proceso específico dependerá del tipo de algoritmo utilizado.

Convencionalmente los algoritmos de *machine learning* se clasifican en procedimientos de aprendizaje supervisado y procedimientos no supervisados, a las que se añade el aprendizaje reforzado y el *deep learning*. En el primer supuesto, la experiencia es transmitida a través de una muestra de datos de entrenamiento (*training set*), que contiene información significativa (*label* o etiqueta) para resolver un problema o realizar un análisis predictivo. Las etiquetas transmiten al algoritmo la información objetiva correcta a partir de ejemplos, que éste utiliza, combina y aplica a nueva información.

A efectos de ejemplificar lo dicho, si quisiéramos clasificar correos electrónicos como legítimos o *spam* (*spam*/*no spam*) utilizando procedimientos de aprendizaje automático supervisado, recurriríamos a un *dataset* con una muestra relevante de correos previamente etiquetados como *spam* o *no spam*. A partir de esta información de entrenamiento, nuestro «aprendiz» deduciría criterios que le permitirían resolver el problema, es decir, catalogar correos electrónicos futuros como legítimos o fraudulentos[10].

8 KELLEHER, J.D., *et. al* (2015), *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*, The MIT Press, p. 3.

9 Una explicación puede verse en MITCHELL, T. (2006), *The Discipline of Machine Learning*, Carnegie Mellon University, CMU-ML-06-108, pp. 1-2. También en RUSSELL, S. Y NORVIG, P., *Ob. Cit*, p. 5-16.

10 El procedimiento, obviamente resulta complejo y no exento de problemas. Una explicación general, puede verse en DADA, E. G., *et al*. (2019) «Machine learning for email spam filtering: review, approaches and open research problems», *Heliyon*, Elsevier, 5(6).

Cómo puede deducirse, en este tipo de aprendizaje la intervención humana juega un papel fundamental al transmitir las etiquetas, y por ende, la respuesta objetiva que guiará el resultado del algoritmo. Por el contrario, en el aprendizaje no supervisado no se transmiten etiquetas o respuestas objetivas. A partir de toda la información, se espera que los algoritmos sean capaces de detectar patrones, correlaciones o categorías. Ejemplos habituales de aprendizaje no supervisado envuelven agrupaciones o *clustering*.

Las aplicaciones del *machine learning* alcanzan un espectro amplio, que van desde la seguridad de la información y la ciberseguridad[11], reconocimiento de voz y procesamiento natural del lenguaje[12], reconocimiento de imágenes y visión computacional[13], resolución de problemas de segmentaciones y clasificación[14], sistemas de recomendación[15], o conducción autónoma,[16] entre muchas otras.

Por su parte, el aprendizaje por refuerzo o aprendizaje semi-supervisado, representa un procedimiento híbrido. De manera similar al aprendizaje no supervisado, el algoritmo no recibe datos de entrenamiento previamente etiquetados, y se espera que a partir de esta información realice inferencias. No obstante, a semejanza con el aprendizaje supervisado, una vez que se obtienen las inferencias, predicciones o clasificaciones, se someten a revisión, bien mediante información etiquetada o contando con intervención humana directa. Esta intervención posterior le proporciona *feedback,* que transmiten experiencia al algoritmo y les permite aprender para resolver el problema asignado. El aprendizaje reforzado ha adquirido relevancia para resolver problemas relacionados con la programación de juegos, robótica, sistemas de recomendación o agentes autónomos (e. g, coches autónomos).

Por otra parte, la definición más aceptada de Big Data es la propuesta por Doug Laney[17] a partir de la idea de 3Vs: Volumen, Variedad y Veloci-

11 Véase BUCZAK, A. L. y GUVEN, E. (2016) «A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection», en *IEEE Communications Surveys and Tutorials*, IEEE, 18(2), pp. 1153 y ss.

12 DENG, L. and LI, X. (2013) «Machine learning paradigms for speech recognition: An overview», en *IEEE Transactions on Audio, Speech and Language Processing*, 21(5), pp. 1060 y ss.

13 MITCHELL, T. (2006), *Ob. Cit.*, p. 2.

14 Al respecto véase, por ejemplo, el trabajo de DADA, E. G. *et al*. (2019), ya citado.

15 PORTUGAL, I., *et al*. (2018) «The use of machine learning algorithms in recommender systems: A systematic review», en *Expert Systems with Applications*, Elsevier, pp. 205 y ss.

16 STILGOE, J. (2018), «Machine learning, social learning and the governance of self-driving cars», en *Social Studies of Science*, 48(1), pp. 25 y ss.

17 LANEY, D. (2001), «3-D Data Management: Controlling Data Volume, Velocity and Variety», en *META Group Research Note*.

dad de los datos. Esto supone la generación de ingentes cantidades de datos, que se producen de manera constante e incremental, y en formatos diversos, dejando atrás los típicos datos estructurados recogidos mediante las tradicionales bases de datos. La digestión y gestión de estos datos, y su aprovechamiento de la mano de técnicas de *machine learning* y *deep learning* para extraer patrones o resolver problemas de amplio espectro, exige acudir a arquitecturas y técnicas específicas[18].

Dicho esto, debemos pasar a explicar la regulación general que se ha venido desarrollando en el entorno europeo, y que tiene como cimientos las disposiciones y lineamientos del Consejo de Europa.

III. El marco normativo del Consejo de Europa respecto a la protección de datos, Big Data y decisiones algorítmicas

En el ámbito del Consejo de Europa, la protección de los datos personales, y por extensión del perfilado y algunas decisiones algorítmicas, encontró en un primer momento, conexión y cobijo en el artículo 8 del Convenio Europeo de Derechos Humanos(CEDH).

Este artículo reconoce el derecho a la intimidad y a la vida familiar como derecho fundamental, que sólo admite restricciones acordes con las leyes internas y en la medida necesaria para salvaguardar objetivos legítimos dignos de protección. Según la jurisprudencia del Tribunal Europeo de Derechos Humanos (TEDH), no se trata de un mero derecho de libertad o no injerencia, sino que conlleva obligaciones positivas para los Estados, que supone, entre otros, la adopción de medidas legislativas dirigidas a garantizar de manera real y efectiva su indemnidad, así como evitar interferencias no admisibles[19]. En el marco de estas obligaciones positivas, encontró engarce la protección de datos personales, cuyo respeto se consideró determinante a efectos de garantizar la protección de la intimidad y la vida familiar. En este sentido, se ha sostenido con acierto que la protección de datos personales representa una manifestación específica del derecho a la vida privada y familiar protegido por el art. 8 CEDH[20].

18 Una explicación sobre este aspecto, puede verse en CABALLERO, R. y MARTÍN, E. (2015), *Las bases del Big Data*, Catarata, Madrid, p. 29 y ss.

19 El *leading case*, en esta materia es el asunto Marckx v. Bélgica, de 13 de junio de 1979, § 31. Esta doctrina ha sido reiteradamente sostenida por el Tribunal. Véase, entre otros, el asunto I contra Finlandia, nº 20511/03, de 17 de julio de 2008, § 35-49.

20 Así, RUIZ MIGUEL, C. (2003), «El derecho a la protección de los datos personales en la Carta de Derechos Fundamentales de la Unión Europea: Análisis crítico», *Revista de Derecho Comunitario Europeo*, Año 7, Núm. 14, enero-abril, p. 11.

Esta primigenia aproximación regulatoria de la protección de datos, imbuida de la evolución del derecho a la privacidad elaborado desde el siglo XIX a partir de las aportaciones de Warren y Brandeis, cedería en los años 60 y 70 ante el avance de la informática y la adopción de un nuevo enfoque, calificado como de segunda generación, que incluye la protección de datos como una dimensión agregada, pero diferenciable, del derecho a la privacidad.

Los aportes e instrumentos posteriores, incluyendo los desarrollados en el marco del Consejo de Europa, supusieron un esfuerzo por reconocer la autonomía del derecho a la protección de datos y su correlativo derecho a la autodeterminación informativa. Desde esta perspectiva, improntada por la sentencia del tribunal constitucional alemán de 1983 en el asunto censo, se devuelve al ciudadano el control de sus datos personales (enfoque de tercera generación), y su protección se refuerza con nuevos derechos conexos dirigidos a garantizar la libre disposición de los datos, con la excepción de aquellos ámbitos cuya afectación se excluye a la disposición individual o se sujeta a especiales restricciones con el fin de proteger los círculos vitales más íntimamente conexos con la dignidad de la persona (enfoque de cuarta generación).

Este enfoque de tercera y cuarta generación, se incorpora al marco del Consejo de Europa a partir de algunos instrumentos, fundamentalmente de carácter principalista y sujetos a vinculación jurídica variada, que pretenden constituir no sólo el marco europeo en este ámbito, sino un referente para el desarrollo de modelos regulatorios y de autorregulación internacionales. Estos instrumentos, que construyen una cascada regulatoria, son los que dan cobertura jurídica a la regulación específica del perfilado ante la ausencia de referencias explicitas en el Convenio de Roma.

1. El Convenio 108 del Consejo de Europa y la normativa complementaria

Los instrumentos que se han ido desarrollando en el consejo de Europa, y que se enmarcan en las normas de tercera y cuarta generación, tienen como punto de referencia el Convenio No. 108 para la protección de las personas con respecto al tratamiento automatizado de datos de carácter

personal[21] y su protocolo adicional[22], precedidos de algunas resoluciones del Consejo de Ministros, dictadas con base en el artículo 8 del CEDH, que mostraron la temprana preocupación del Consejo de Europa por la protección y manejo adecuado de los datos, tanto por el sector público[23] como por agentes privados[24].

El convenio es un instrumento específico y sistemático, cuyo objetivo es proteger el derecho a la privacidad y la autodeterminación informativa, frente a los riesgos que avizoraban los avances de las ciencias informáticas. Su especial trascendencia radica en que, a la fecha, es el único instrumento internacional vinculante en materia de protección de datos, y representa una propuesta de armonización y norma modelo de escala global[25]. Además, su influencia a nivel europeo no se ha reducido exclusivamente a los confines del Consejo de Europa, sino que ha servido de fuente de inspiración a la normativa comunitaria de protección de datos de 1995[26], y dio paso a la adhesión de las Comunidades Europeas en 1999. A partir de allí, se inició un continuo proceso de acercamiento, cooperación y armonización entre ambas organizaciones, lo que deja constancia del impacto real que el instrumento ha tenido en diversas jurisdicciones y niveles normativos.

21 Convenio número 108 el Consejo de Europa, de 28 de enero de 1981, para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal. El convenio, fue aprobado y abierto para su ratificación el 28 de enero de 1981, y entró en vigor el 1 de octubre de 1985. A fecha de hoy, ya ha sido ratificado por los 47 países miembros del consejo de Europa

22 Protocolo adicional al convenio para la protección de las personas con respecto al tratamiento automatizado de datos de carácter Personal relativo a la transferencia de datos, en lo que respecta a las autoridades de control y flujos transfronterizos de Datos, CETS, No. 181, 2001.

23 Resolución (74) 29 relativa a la protección de la vida privada de las personas físicas respecto de los bancos de datos electrónicos en el sector público, de 20 de septiembre de 1974.

24 Resolución (73) 22 relativa a la protección de la vida privada de las personas físicas en relación con los bancos de datos electrónicos en el sector privado, de 26 de septiembre de 1973.

25 El convenio es fruto del consenso y participación no sólo de los Estados Parte, sino de países occidentales no miembros como Canadá, Estados Unidos y Japón. Está abierto a adhesión por terceros países, aunque bajo un estricto proceso de invitación, al que se han sumado Mauritania, Senegal, Uruguay y Tunes, mientras que otros aún no han completado el proceso.

26 Directiva 95/46/CEE, del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección que las personas físicas en lo que respecta al tratamiento de datos personales y a la libre Circulación de estos datos.

El convenio es jurídicamente vinculante, si bien es cierto que sustantivamente presenta una densidad normativa bastante escasa[27]. No obstante recoge, principios y garantías aplicables a autoridades públicas y actores privados que orientan la recopilación y tratamiento de datos personales (calidad de los datos[28] y procesamiento de datos sensibles[29]), así como la necesidad de informar a las personas afectadas sobre sus derechos de acceso y rectificación[30], todos ellos verdaderas claves de bóveda y elementos reticulares de la regulación del perfilado, tanto a nivel comunitario como del propio Consejo de Europa.

Con fundamento en el Convenio se ha suscrito un protocolo adicional, que incorpora salvaguardas adicionales, y el Comité de Ministros del Consejo de Europa ha adoptado diversas recomendaciones dirigidas a sectores específicos[31].

Estas recomendaciones, pese a no tener carácter vinculante, constituyen parámetros de referencia para la interpretación del derecho a la pri-

27 La elección de un modelo basado en principios, es razonable sostener, obedece a la naturaleza misma del Consejo de Europa, su composición y fines, por una parte, y al señalado carácter abierto de la Convención 108 y su interés en promoverlo como un instrumento de armonización internacional, por otra.
El Consejo de Europa, a diferencia -por ejemplo- de la Unión Europea, tiene una composición más heterogénea y no ha vivido un proceso armonizador y de cohesión política, económica y social. Las diferencias entre sus miembros son mucho más pronunciadas que las que existen en la Unión Europea, y su alcance territorial mucho más amplio. De ahí que su forma de actuación sea la propia de una organización internacional que busca fijar estándares mínimos aceptables por todos los Estados parte, y reduzca a la expresión necesaria las muestras jurídicamente vinculantes de sus instrumentos.
Esto, en el caso del Convenio 108, ha facilitado que sea tenido como una referencia en otras jurisdicciones y un primigenio estándar internacional de protección de datos, dado su carácter principalista relativamente simple, conciso y tecnológicamente neutro. Véase POLAKIEWICZ, J. (2011), «Convention 108 as a global privacy standard?», *International Data Protection Conference*, Budapest, p. 2

28 En cuanto a la calidad de los datos el convenio establece la obligación de que los datos almacenados den cumplimiento al principio de proporcionalidad, esto es, deben ser adecuados y alcanzar aquello que sea necesario para los fines legítimos perseguidos. Asimismo, tanto la recogida de datos como su tratamiento deben ser lícitos y legítimos, además de conservarse durante el período estrictamente necesario para los fines previstos.

29 Respecto a los datos considerados sensibles (origen étnico, Ideología política, datos relacionados con la salud, orientación sexual, creencias religiosas, antecedentes penales), sólo permite su recopilación y tratamiento si se brindan garantías y salvaguardas adecuadas de conformidad con los estándares de protección previstos en el convenio.

30 Este aspecto, es uno de los que mayor impacto ha tenido, ya que más adelante se configurarían, con la evolución del régimen europeo de protección de datos, como los derechos ARCO, es decir, el derecho de las personas físicas al acceso y conocimiento de los datos que sobre ellas se conserven, así como, el derecho a obtener su rectificación.

31 Así, se han dictado recomendaciones sobre la recolección y procesamiento de datos en el sector de los seguros, Recomendación Rec(2002)9 sobre la protección de datos de carácter personal recopilados y tratados con fines de seguro, 18 de septiembre de 2002., sobre la protección de datos a efectos estadísticos, Recomendación del Consejo de Ministros R (97)18 relativo a la protección de datos personales recopilados y procesados con fines estadísticos, de 30 de septiembre de 1997, y para fines de investigación científica y estadística Recomendación del Consejo de Ministros R (97)18 relativa a la protección de datos personales utilizados con fines de investigación científica y estadística, sobre la protección de datos con fines de marketing directo Recomendación del Consejo de Ministros R (85)20 sobre la protección de datos personales utilizados con fines de marketing directo, entre otras.

vacidad y la protección de datos. En otras palabras, desarrollan en sectores concretos los principios previstos en el convenio, y se asimilan a instrumentos intermedios de dirección que brindan mayor transparencia y seguridad jurídica a los operadores y personas afectadas. De estas directrices, nos resultan de especial interés las que regulan el tratamiento automatizado de datos y el perfilado, a las que hemos hecho referencia en el apartado anterior, y a cuya explicación nos remitimos, por lo que nos resta por apuntar las que recogen la respuesta regulatoria del Consejo respecto al tratamiento de datos utilizando técnicas de *big data*. Esto es, los lineamientos sobre la protección de las personas respecto al procesamiento de información personal en un mundo de Big Data[32].

Los lineamientos tienen como objetivo establecer un marco general de aplicación de políticas y medidas que hagan efectivos los principios y demás previsiones del Convenio 108 frente a los retos del *big data*. Destacan, además, por ser el primer instrumento internacional dirigido a regularlo[33], y a diferencia del resto de recomendaciones dirigidas a un ámbito de aplicación sectorial concreto, adoptan un enfoque transversal pues se centran en una tecnología específica.

Además de lo señalado, desde nuestra perspectiva dos aspectos son destacables. Por una parte, el instrumento supera la visión técnica que ha prevalecido en la definición canónica de esta tecnología y coloca como centro gravitacional sus posibles disfunciones y riesgos, así como la inoperatividad de la aproximación regulatoria que informa los instrumentos previamente adoptados en materia de protección de datos. En este sentido, adopta un concepto mucho más amplio, que destaca el componente predictivo y su uso en la adopción de decisiones con impacto sobre los derechos y patrimonio jurídico de las personas, lo que encuentra franca conexión con el perfilado. La ampliación del enfoque es fácilmente justificable: el avance tecnológico obsede y rebasa el escenario conforme al cuál se diseñó la normativa previa. El limitado nivel de complejidad del análisis de datos, así como la capacidad de los individuos para entender sus resultados, que justificaron el alcance de la regulación prevista en el convenio 108, ya no pueden sostenerse ante el desarrollo de las modernas tecnologías que hemos descrito. La actuación de complejos algoritmos que permiten obtener inferencias ininteligibles para los usuarios a partir de las ingentes cantidades de datos que puede procesarse a través del Big Data, se aparta radicalmente

32 T-PD (2017)01, aprobados por el Comité Consultivo del Convenio 108.

33 MANTELERO, A. (2017), «Regulating big data. The guidelines of the Council of Europe in the context of the European data protection framework», *Computer Law and Security Review*, 33(5), p. 591.

del escenario regulatorio al que pretendía dar respuesta el convenio 108, ayuno de estas valoraciones.

Por otra parte, los lineamientos refuerzan el enfoque basado en derechos. En este sentido los lineamientos, además de reiterar el crucial y estructural valor de los derechos humanos y la dignidad de la persona, establecen algunos principios generales y destacan la importancia de promover a nivel interno un marco de análisis y políticas preventivas respecto a los posibles riesgos derivados del uso del *big data*. En otras palabras, recoge un enfoque preventivo que tiene como referente el principio de precaución. Centrándonos en el perfilado, exige que se garantice un cierto nivel de transparencia e información del sujeto afectado por estas decisiones, que explique de manera suficiente el razonamiento y la lógica que subyace.

El Convenio 108 y las directrices mencionadas, a día de hoy cierran el marco del Consejo de Europa en materia de perfilado y decisiones automatizadas, y complementa el puzle regulatorio cuyas piezas significativas restantes las encontramos en la normativa comunitaria, especialmente en el nuevo Reglamento General de Protección de Datos que afina las escasas orientaciones previstas por la regulación anterior.

2. Lineamientos sobre procesamiento de información en un mundo de Big Data

Los lineamientos sobre la protección de las personas respecto al procesamiento de información personal en un mundo de Big Data (T-PD(2017)01), aprobados por el Comité Consultivo del Convenio 108, tienen como objetivo establecer un marco general de aplicación de políticas y medidas que hagan efectivos los principios y previsiones del Convenio 108, en el contexto del big data. Las directrices, comparten la misma naturaleza que el resto de recomendaciones dictadas en el marco del Convenio 108, por lo que no resultan legalmente vinculantes, aunque representan lineamientos de interpretación. Destacan además por ser el primer instrumento internacional dirigido a regular el *big data*[34], y a diferencia de las recomendaciones antes citadas del Consejo con un ámbito de aplicación sectorial concreto, se centran en una tecnología específica, por lo tanto con ámbito trasversal sujeta a complitud por la normativa y directrices sectoriales que resulten aplicables.

34 MANTELERO, A., *Ob. Cit.*, p. 591.

Los lineamientos recogen una definición de *big data* que asume se aleja de la concepción tradicional. Adoptan un concepto mucho más amplio, que destaca el componente predictivo y su uso en la adopción de decisiones con impacto sobre los derechos y patrimonio jurídico de las personas[35]. La ampliación del enfoque se justifica en el avance tecnológico que deja en desuso el escenario conforme al cuál se diseñó la normativa previa. El limitado nivel de complejidad del análisis de datos, así como la capacidad de los individuos para entender sus resultados, que justificaron el alcance de la regulación prevista en el Convenio 108, ya no pueden sostenerse ante el desarrollo de las modernas tecnologías que ya hemos descrito. La actuación de complejos algoritmos que permiten obtener inferencias ininteligibles para los usuarios a partir de las ingentes cantidades de datos que puede procesarse a través del *big data*,[36] se aparta radicalmente del escenario regulatorio previo.

En este sentido los lineamientos, además de reiterar el crucial y estructural valor de los derechos humanos y la dignidad de la persona, establecen algunos principios generales y destacan la importancia de promover a nivel interno un marco de análisis y políticas preventivas respecto a los posibles derivados del uso del *big data* basado. En otras palabras, recoge un enfoque preventivo que tiene como referente el principio de precaución.

Por otra parte, y en cuanto a los aspectos más relacionados con las decisiones automáticas y posibles discriminaciones algorítmicas, promueve —no de forma tan directa como podría— un enfoque basado en derechos, sugiriendo que se garantice la transparencia e información del sujeto afectado por estas decisiones, y explicando de manera suficiente el razonamiento y la lógica de las decisiones o resultados automáticos.

Un elemento que resulta novedoso en la construcción de los lineamientos es la necesidad de preservar para los sujetos controladores la autonomía humana en la valoración de las decisiones automatizadas, con la correlativa libertad para apartarse de ellas. También resulta innovadora la propuesta de inversión de la carga de la prueba en casos de posibles discriminaciones, directas o indirectas, siendo los sujetos controladores o responsables del procesamiento los que deberían probar la ausencia de un trato diferenciado injustificado, así como la necesidad de establecer un derecho a recurrir estas decisiones ante una autoridad a la que se le hayan otorgado competencias.

35 Al respecto véase el apartado 3 de las directrices.

36 MANTELERO, A., *Ob. Cit.*, p. 585.

Finalmente, un aspecto no baladí, que aunque no sea objeto de este trabajo justificaría posteriores análisis, es que los lineamientos destacan la importancia de incorporar una dimensión colectiva al derecho a la privacidad y la autodeterminación informativa en el ámbito del *big data*, que podemos extender sin mayores impedimentos a las decisiones algorítmicas y al perfilado. Esta dimensión colectiva[37], podría llevar a reconsiderar el modelo vigente de protección de datos que se mantiene en el Derecho comunitario y en el Reglamento General de Protección de datos, y los lineamientos constituyen uno de los primeros instrumentos en avizorarlo.

Los lineamientos, que a día de hoy cierran el marco del Consejo de Europa[38] en materia de perfilado y decisiones automatizadas, complementa el marco regulatorio cuyas piezas significativas restantes encontramos en la normativa comunitaria, especialmente en el nuevo Reglamento General de Protección de Datos que afina las escasas orientaciones previstas por la regulación anterior.

IV. Regulación de las decisiones algorítmicas y de perfilado en el Reglamento General de Protección de Datos

A nivel europeo, la directiva de protección de datos de 1995[39] no recoge ninguna previsión acerca del perfilado y la adopción de decisiones basadas en él, aunque si se ocupa de las decisiones automatizadas[40].

La primera institución en ocuparse de esta cuestión con profundidad fue el Consejo de Europa, que aprobó en 2010 la recomendación sobre la protección de las personas con respecto al procesamiento electrónico de datos en el ámbito del *profiling*, Recomendación CM/Rec(2010)13. Pese al transcurso de casi una década desde su aprobación, la recomendación aún representa uno de los instrumentos más avanzados respecto a la regulación del perfilado. Su influencia ha sido determinante en ins-

37 La dimensión colectiva del derecho a la privacidad como consecuencias de las nuevas tecnologías ha sido expuesta, entre otros trabajos, en TAYLOR, L., *et.al*, (eds.), (2017), *Group Privacy New Challenges of Data Technologies*, Springer, Dordrecht.

38 Cabe señalar que actualmente se está negociando la modernización del Convenio 108, con el objetivo de dar respuesta a los retos que conllevan las nuevas tecnologías, y convertirlo en un marco multilateral y flexible, que sintetice los diferentes modelos regulatorios y permite el intercambio transfronterizo de datos en un entorno de seguridad y protección de los derechos de los ciudadanos.

39 Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos.

40 Un análisis reciente en PALMA ORTIGOSA, A. (2019), «Decisiones automatizadas en el RGPD. El uso de algoritmos en el contexto de la protección de datos», *Revista General de Derecho Administrativo*, Núm. 50.

trumentos aprobados con posterioridad por la Unión Europea, especialmente con miras a *juridificar* los contornos de una técnica que nació y se desarrolló a extrarradios del mundo del Derecho.

Así pues, la recomendación aborda una definición de perfilado diferenciándolo a su vez del perfil, a pesar de su estrecha vinculación. El primero, vendría a ser una técnica de procesamiento automático de datos que consiste en aplicar un perfil a un individuo con el objetivo bien de tomar decisiones sobre esta persona, o bien analizar o predecir sus preferencias, comportamientos o actitudes.[41] El perfil, por su parte, hace referencia a una serie de datos que caracterizan una categoría de individuos y que se pretenden aplicar a un individuo específico[42].

Como se ve, el Consejo adopta una definición de naturaleza funcional, cuya construcción descansa en las distintas fases del proceso de perfilado. A tales efectos, se entiende por tal un proceso trifásico que incluye una fase de recolección de datos; otra de análisis automático de éstos para identificar correlaciones; y, finalmente, la aplicación de esas correlaciones a individuos específicos deduciendo así su posible comportamiento pasado, presente o futuro.

Esta definición comentada, eminentemente funcional, constituye una referencia para el marco comunitario, y ha servido de inspiración al Reglamento General de Protección de Datos, así como a la Directiva sobre protección de los datos personales tratados a efectos policiales y judiciales,[43] aunque ambos instrumentos recogen una aproximación técnicamente más completa y autoexplicativa. El RGPD define el perfilado, o de manera más certera la elaboración de perfiles, como «toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales de una persona física, en particular para analizar o predecir aspectos relativos al rendimiento profesional, situación económica, salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos de dicha persona física» (art. 4.4).

Desde el punto de vista de la doctrina, Uno de los primeros intentos por adoptar una definición comprensiva de esta técnica, fue realizado por Clarke hace un cuarto de siglo. Este autor, referencia habitual de quienes se han ocupado del tema[44], concibe el perfilado como una técnica de vigilancia de datos, a partir de la cual una serie de características de un grupo de personas es inferida de su comportamiento pasado, para posteriormente buscar individuos que encajen en ese mismo grupo y características[45]. En términos similares, Hildebrandt propuso una definición funcional, que se considera canónica en la materia, y que concibe el perfilado como el proceso de «descubrimiento» de correlaciones en los datos que pueden ser utilizados para identificar o representar a sujetos humanos o no humanos (individuos o grupos), y/o la aplicación de perfiles para individualizar y representar a un sujeto específico o identificarlo como miembro de un grupo[46].

Un concepto con estrecha conexión al perfilado, y del que puede predicarse un cierto solapamiento en su definición y aplicación práctica, son las decisiones automáticas o algorítmicas. La directiva europea de 1995 apuntaló en su artículo 22 un escaso marco jurídico, reconociendo el derecho a no verse sometido a decisiones individuales automatizadas que conlleven efectos jurídicos o afecten de manera significativa similar.

El Reglamento general de protección de datos recoge en buena medida, también en su artículo 22, la letra de la Directiva 95/46/CE, aunque armonizando de manera más intensa su aplicación a nivel comunitario. El dispositivo coloca implícitamente en el núcleo de la regulación el principio de autodeterminación informativa y el control de los datos.

Acoge un enfoque con un marcado carácter garantista, puesto que refleja tanto una prohibición general de decisiones automatizadas, como un derecho a oponerse a este tipo de decisiones si producen efectos significativos o le afectan de manera particular. El reconocimiento del derecho se recoge explícitamente en el art. 22 del RGPD, así como en el considerando 71 de su exposición de motivos.

Ahora bien, este enfoque garantista se refleja y refuerza con la implícita adopción de un enfoque basado en derechos como aproximación para responder a los retos que plantea el perfilado, y las decisiones automatizadas. Esto refleja una extensión más del derecho a la autodeterminación informativa, que se integra por las facultades y derechos que se reconocen a los afectados o sujetos activos para proteger sus datos y su

41 Apéndice 1.e de la Recomendación CM/Rec(2010)13.

42 Apéndice 1.d de la Recomendación CM/Rec(2010)13.

43 Véase art. 3.4 de la Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos.

44 Así, entre otros, BOSCO, F., *et. al* (2015), «Profiling Technologies and Fundamental Rights. An Introduction», en CREEMERS, N., et.al (ed.), *Profiling Technologies in Practice. Applications and Impact on Fundamental Rights and Values*, WLP, Oisterwijt, p. 8.

45 CLARKE, R. (1993), «Profiling: A Hidden Challenge to the Regulation of Data Surveillance», en *Journal of Law, Information and Science*, Volumen 4, Núm. 2.

46 HILDEBRANT, M. (2008), «Defining Profiling: A New Typo of Knowledge? », en HILDEBRANT M., y GUTWIRTH, S., *Profiling the European Citizen: Cross-Disciplinary Perspectives*, Springer, Dordrecht, p. 19.

posición jurídica[47], y que en el Derecho interno el Tribunal Constitucional ya reconoció desde su sentencia 290/2000[48].

Algunos de estos derechos que recoge el RGPD suponen un refuerzo o la aplicación de especialidades de derechos consolidados tradicionalmente conocidos como derechos ARCO, sobre los que el extenso tratamiento ya realizado por la doctrina hace ocioso y reiterativo su tratamiento en este trabajo[49]. Si conviene, por el contrario, que dediquemos el análisis a derechos y salvaguardas específicas recogidas por el Reglamento y que han generado cierta discusión respecto a su alcance y virtualidad por parte de la doctrina.

Así, el RGPD recoge el derecho a oponerse a decisiones automáticas y de perfilado, como un derecho autónomo y con perfil propio que se desgaja de los conocidos derechos ARCO. El artículo 21 del Reglamento lo recoge, pero construido sobre cierta complejidad que requiere apuntalar alguna aclaración.

La norma exige una ponderación entre los posibles derechos e intereses del afectado, y otros motivos legítimos imperiosos, pero sin llegar a definir o aventurar criterios para valorarlos. Las directrices del WP29 sobre decisiones automatizadas y elaboración de perfiles, refiere supuestos donde el perfilado, por ejemplo, pueda considerarse beneficioso para la sociedad al aplicarse para predecir la propagación de enfermedades contagiosas[50], pero más allá de esto, encontramos una cierta orfandad de criterios.

No explicita tampoco a quien corresponde realizar la ponderación, ni en qué medida le está permitido al interesado participar. No obstante, puede concluirse, que la valoración se deja en manos del responsable del tratamiento a tenor de la letra del artículo 21.1 del Reglamento. Ahora, si el resultado de la ponderación arroja la prevalencia del motivo imperioso sobre los derechos e intereses del afectado, puede procederse con el perfilado y no procede activar, como consecuencia, los derechos de supresión o limitación en la finalidad justificados en la oposición. Caso contrario, debe suspenderse la elaboración del perfilado o su aplicación.

Cabe señalar que el RGPD contempla como caso especial el supuesto del tratamiento dirigido al perfilado con fines de mercadotecnia, y recoge una solución radical prohibiendo en caso de oposición la elaboración o aplicación del perfilado, y en caso de petición del afectado, la supresión de sus datos.

Por otra parte, el Reglamento recoge también con carácter propio y alcance específico el novedoso derecho a la intervención humana. A este respecto, uno de los aspectos críticos radica en determinar el alcance y contenido de dicha intervención.

En cualquier caso, lo que pretende este dispositivo en que la participación cognitiva de un ser humano vinculado con la decisión no sea vacua. De ahí que tenga sentido señalar que intervenciones accesorias o instrumentales no resultan suficientes para dar cumplimiento a las exigencias del Reglamento. En otras palabras, debe ser suficientemente relevante y debe atribuirse capacidad y autorización para revertir, modificar, o transformar la decisión adoptada algorítmicamente[51]. Esto exige, por tanto, intervenciones de calado y no meramente cosméticas, así como explicar las razones por las que adoptan las decisiones, en términos suficientemente claros para que el afectado lo comprenda, teniendo como sujeto receptor una persona media sin conocimientos técnicos o jurídicos especiales.

Finalmente, y en conexión con esto último, el RGPD establece en el art. 22.3 el derecho a la explicación, especialmente significativo para nuestro objeto de estudio. Esta salvaguarda pretende dar respuesta a la opacidad y escasa legibilidad de los análisis obtenidos mediante la aplicación de inteligencia artificial, por ejemplo para realizar perfilados o adoptar decisiones automatizadas, y que algunos autores asocian a la existencia de cajas negras[52].

La idea de caja negra hace referencia no sólo a la opacidad de la información, sino también a su legibilidad, esto es, la capacidad de que la información y los resultados puedan ser interpretados y aprehendidos por los usuarios afectados. Por tanto, no basta con garantizar la transparencia del proceso o el acceso a los resultados, puesto que por razones técnicas los datos son incomprensibles, o las implicaciones de los resultados del tratamiento sólo están al alcance de expertos. De ahí que el RGPD exija a los responsables del tratamiento que transmitan una explicación significativa sobre la lógica aplicada por los algoritmos, así como sobre la importancia y consecuencias que deriven de las decisiones.

Ahora bien, la existencia del derecho no ha estado exento de controversias y ha sido cuestionado desde la doctrina, donde encontramos

47 En este sentido en términos generales, LUCAS MURILLO, P. (1990), *El derecho a la autodeterminación informativa*, Tecnos, Madrid, p. 185.

48 STS 290/2000, de 30 de noviembre.

49 Al respecto, entre otros, véase SERRANO PÉREZ, M.M. (2003), *El derecho fundamental a la protección de datos. Derecho español y* comparado, Thomson-Civitas, Cizur Menor, pp. 343 y ss.; HERNÁNDEZ LÓPEZ, J.M. (2013), *El Derecho a la protección de Datos Personales en la Doctrina del Tribunal Constitucional*, Thomson Reuters-Aranzadi, Cizur Menor, pp. 74 y ss.

50 WP29, Directrices sobre decisiones automáticas y elaboración de perfiles, p. 20.

51 En este sentido, WP29, Directrices sobre decisiones automatizadas y elaboración de perfiles, p. 30.

52 Sobre la idea de cajas negras, y la problemática que plantean, véase PASQUALE, F. (2015), *The Black Box Society. The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge.

voces a favor[53] y en contra[54] de su reconocimiento. Desde nuestra perspectiva, es razonable defender y reconocer este derecho, que —entendemos— deriva de la exigencia del Reglamento a trasmitir información significativa sobre la lógica aplicada para adoptar decisiones automáticas y de perfilado.

Es cierto que la regulación en este aspecto dista mucho de ser precisa y acabada[55], y que existen aspectos cuya delimitación se deja en manos de las autoridades administrativas y los tribunales, pero consideramos acertado recurrir a una interpretación sistemática del RGPD que acoge un estándar garantista.

En línea con esto, garantizar el acceso a la explicación de cómo las decisiones que afectan a los ciudadanos han sido tomadas, especialmente si estamos ante procesos opacos, contribuye a garantizar el Estado de Derecho. La negación de esta información, no sólo reduce las posibilidades de defensa de los propios intereses y derechos, sino que imposibilita su ejercicio dado que dificulta conocer el razonamiento sobre el cuál se han adoptado las decisiones que les afectan. No exige mayor argumentación sostener que la posibilidad de defenderse frente a posibles afectaciones descansa en el acceso previo a la explicación de sus razonamientos y motivaciones.

Habiendo abordado en términos generales la regulación de las decisiones automáticas y de perfilado, procede ahora analizar el caso particular y la problemática jurídica asociada al perfilado con fines ideológicos.

V. El perfilado político o con fines ideológicos: aproximación a su regulación y problemática ético-jurídica

El perfilado ideológico tiene como objetivo ganar en influencia estratégica e inclinar el voto de los electores. Supone el desarrollo de una serie de fases, que incluyen la recolección de información personal de los votantes, con el objetivo de enviar mensajes personalizados[56].

53 Hacen referencia a la existencia de este derecho GOODMAN, B. y FLAXMAN, S. (2017), «European Union regulations on algorithmic Decision-making and a right to explanation», *AI Magazine*, Vol. 38, Núm. 3.

54 Al respecto, véase WATCHER. S., MITTELSTADT, B. y FLORIDI, L. (2017), «Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation», en *International Data Privacy Law*, Vol. 7, Núm. 2, pp. 76 y ss. Para estos autores, no es posible derivar la existencia de tal derecho desde las bases que identifican en el RGPD, principalmente el derecho de acceso a la información.

55 Comparten esta posición EDWARDS, L. y VEALE, M. (2018), «Enslaving the algorithm: from a 'right to an explanation' to a 'right to better decisions'?», *IEEE Security & Privacy*, Vol. 16, Núm. 3, p. 48.

56 Así, ZUIDERVEEN B., F. J. (2018), «Online Political Microtargeting: Promises and Threats for Democracy», en *Utrecht Law Review*, Vol. 14, Núm. 1, p. 82.

El Big Data y la analítica avanzada de datos, permite afinar en la perfilación de usuarios e incorporar elementos o puntos de información obtenidos a partir de inferencias derivadas de puro poder computacional. El resultado, o al menos su reclamo públicamente anunciado, es alcanzar una elevadísima precisión y una eficacia no conocida previamente al depurar audiencias de nicho, para activar o asegurar la base de votantes de un partido, o bien personalizar mensajes dirigidos a persuadir e inclinar la decisión de votantes indecisos.

Ahora bien, ni el *microtargeting* ni los intentos de segmentación de mensajes, resultan ajenos a las campañas electorales de los últimos lustros. La metodología que se aplica se ha desarrollado desde la ciencia y la comunicación política, incorporando técnicas de *machine learning* o *deep learning*, añadiendo sofisticación y teórica efectividad respecto a la propaganda electoral generalizada. La acumulación masiva y la permeabilidad social respecto a la cesión de datos, facilitaría este cambio cualitativo y derivaría en una ruptura respecto a la predicción tradicional basada en modelos diseñados a partir de variables generales (e. g, edad, género o situación socioeconómica).

El proceso de perfilado y microsegmentación de mensajes no está exento de complejidad, y requiere del desarrollo de tres fases. En primer lugar, la construcción de una infraestructura de datos con información de los posibles electores; en segundo término, un proceso de perfilado de usuarios y predicción basado en modelos estadísticos avanzados; y finalmente, una fase centrada en la comunicación con los usuarios. Las dos primeras fases, se centran en el *target* de los electores que interesan de cara a la campaña, y la última en la intervención mediante el envío de mensajes específicos que animen o desactiven el voto.

Ahora bien, desde el punto de vista regulatorio, la respuesta al perfilado ideológico se ha apuntalado desde el derecho a la protección de datos, a nivel comunitario, parcialmente por el RGPD, y a nivel interno, mediante la reforma de la Ley Orgánica de Régimen Electoral, posteriormente declarada inconstitucional por el Tribunal Constitucional.

Respecto a la norma europea, debe señalarse que no reguló de manera exhaustiva el perfilado ideológico, aunque estableció el marco general del perfilado al que hicimos referencia en el apartado anterior. Es llamativa su decisión de no prohibirla expresamente, delegando en el legislador interno la determinación de la necesidad y alcance de un marco específico, así como las condiciones admisibles para su aplicación.

Interesa reseñar, en línea con esto, que el reglamento comunitario no proscribe la recopilación ni el tratamiento de datos personales de naturaleza política realizado por partidos o asociados políticas, aunque lo somete a un régimen particular que extrema las salvaguardas y garantías.

En este sentido, se establece una excepción a la prohibición general de tratamiento de datos de categorías especiales (considerando 56 RGPD), que vincula a una necesidad exigible para el funcionamiento del sistema democrático, sujeta además a autorización normativa de los Estados miembros[57]. De esto se colide que la recopilación y tratamiento de datos vinculados al perfilado ideológico debe radicarse necesariamente en una base jurídica legítima que respete los principios de necesidad y proporcionalidad[58], como exige el art. 8 del CEDH. Aunque, esta restricción puede levantarse si se otorga consentimiento del afectado, encuentra cobijo en el cumplimiento de una misión legal o se sustenta en razones de interés público.

De las bases jurídicas recogidas en el art. 6 RGPD[59], razonablemente puede concluirse que sólo resultan admisibles el tratamiento con fines de perfilado ideológico basado en el consentimiento de los electores, o que se justifique en un tratamiento necesario para la realización de una misión de interés público. El legislador orgánico español acogió como opción para desarrollar la habilitación del RGPD, en esta última, contando con la mediación de la Ley Orgánica 3/2018, de 6 de diciembre, de Protección de Datos y Garantía de los Derechos Digitales (LOPDGDD), que reformó la Ley Orgánica de Régimen Electoral General (LOREG), incorporando el art. 58.bis.

Este artículo fue parcialmente objeto de recurso de inconstitucionalidad promovido por el Defensor del Pueblo. El Alto Comisionado de las Cortes sostuvo, entre otros razonamientos, la vulneración del contenido esencial del derecho a la protección de datos personales (art. 18.4), en conexión con el derecho a la libertad ideológica (art. 16) y el derecho a la participación política (art. 23.1), debido a que el Legislador nacional no concretó las garantías adecuadas para el tratamiento, como exige la doctrina del Tribunal Constitucional.

57 La excepción se justifica por el Legislador comunitario, en la concurrencia de un interés público que legitima suficientemente su tratamiento. En este sentido, informe N/Ref 210070/2018, Gabinete Jurídico de la Agencia Española de Protección de Datos, p. 6.

58 Como es sabido, la construcción del derecho a la protección de datos en conexión con el derecho a la privacidad, se cimienta sobre la inadmisión de interferencias salvo que sean exigidas por la ley o se consideren necesarias en el marco de una sociedad democrática Esto dio pie a la exigencia de base jurídica para el tratamiento de los datos, que actualmente recoge en el art. 6 del RGPD. Han abordado este aspecto desde diversas perspectivas, entre otros, LUCAS MURILLO, P. (1990), *Ob. Cit.*; SERRANO PÉREZ, M. M., (2003), *Ob. Cit.*

59 De manera tasada, el tratamiento sólo se considera lícito si existe consentimiento para el tratamiento para fines específicos (apdo. 1.a); se desarrolla en el marco de un contrato (apdo. 1.b); es necesario para el cumplimiento de una obligación legal (apdo. 1.c); se realice para proteger intereses vitales del interesado o de otra persona física (apdo. 1.d); sea necesario para el cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos conferidos (apdo. 1.e); o sea necesaria para satisfacer intereses legítimos perseguidos por el responsable del tratamiento o un tercero, siempre que no prevalezcan otros intereses, derechos o libertades fundamentales del interesado (apdo. 1.f).

El alto tribunal acogió parcialmente el criterio del Defensor del Pueblo, y declaró la inconstitucionalidad del art. 58.1 bis de la LOREG. Pero no se pronuncia sobre la legalidad del perfilado ideológico o la microsegmentación, ni la proscribe. Por tanto, es procedente abordar algunos aspectos a efectos de contribuir en una futura reforma de esta figura, especialmente respecto a su admisibilidad.

En primer lugar, debemos hacer referencia —con la brevedad del caso— al posible sujeto legitimado para el tratamiento y su posición en el marco del Estado de partidos. La regulación de los partidos políticos, así como el especial papel que atribuye la propia constitución, ha sido objeto de estudio por la doctrina[60], depurado por la jurisprudencia del Tribunal Constitucional[61], por lo que poco cabe añadir.

Dicha jurisprudencia reconoce su especial relevancia constitucional (STC 18/1984, de 7 de febrero, FJ 3), que asimila a una cierta función pública a la que anuda algunos privilegios (STC 3/1981, de 2 de febrero, FJ 2). Estas funciones, de clara trascendencia política, las apuntala con el máximo rango el art. 6 CE (expresión del pluralismo político, concurrencia en la formación y manifestación de la voluntad popular, e instrumento de participación política), en sintonía con el Consejo de Europa y la Comisión de Venecia[62].

La proyección práctica y trascendental de estas funciones en el ejercicio democrático fue tempranamente recogida por la mejor doctrina, cuya mediación es menester para actualizar los principios democráticos[63]. En este sentido, ejercen un importante rol de movilización del electorado, que puede encontrar conexión con el perfilado ideológico.

Desde esta perspectiva, la microsegmentación electoral puede resultar positiva para el sistema democrático en la medida que incrementaría la competencia electoral, estimulando a la vez la participación electoral o, al menos, contribuyendo a reducir la apatía de los votantes. Por otra parte, la racionalidad democrática exige que los electores dispongan de información para ejercer el derecho al sufragio activo. Se espera que los partidos políticos contribuyan con los medios de comunicación a poner en disposición de los electores información sobre los asuntos públicos,

60 Entre otros, VIRGALA FORURIA, E. (2017), *La regulación jurídica de los partidos políticos*, Ediciones Olejnik, Santiago de Chile; MARTÍNEZ CUEVAS, M. D. (2006), *Régimen jurídico de los partidos políticos*, Marcial Pons, Madrid.

61 Véase PRESNO LINERA, M. A. (2000), *Los partidos políticos en el sistema constitucional español. Prontuario de jurisprudencia constitucional: 1980-1999*, Aranzadi, Pamplona.

62 Al respecto véase, OSCE/ODIHR y Comisión de Venecia, Guidelines On Political Party Regulation (2010), apdo. 10.

63 GARCÍA-PELAYO, M. (1986), *El Estado de Partidos*, Alianza, Madrid, p. 74.

así como su oferta de soluciones o alternativas legitimas. Más problemática resulta la microsegmentación de mensajes. De partida asumimos que esta microsegmentación, no debe considerarse *per se* perniciosa. No obstante, desde una perspectiva objetiva la segmentación de votantes y, en consecuencia, el envío de mensajes ajustados a sus intereses en un entorno de extremo ruido mediático, puede ofrecer algunas ventajas. Desde esta perspectiva, la microsegmentación puede amplificar efectos positivos[64]. Así, permite enviar mensajes de la agenda política que puedan resultar de mayor interés para cada segmento de votantes. Aunado a esto, puede suponer una ventaja para ganar en diversidad en el mercado de ideas electoral. Por tanto, acompasando las legítimas funciones de los partidos políticos, vertería ventajas para el proceso democrático y contribuiría también a facilitar la información a la que acceden los electores en el mercado de las ideas, y en último término fomentaría la pluralidad y el diálogo democrático.

No obstante, en sentido opuesto, deben reconocerse posibles patologías. En primer lugar, el *microtargeting* puede contribuir a la manipulación electoral. Si a esto sumamos la posibilidad de minimizar la transparencia y el debate, la posibilidad de rebatir la veracidad de los contenidos y el debate democrático se desvanece. Por otra parte, plantea el riesgo de agudizar infrarrepresentaciones existentes[65], así como contribuir a la excesiva fragmentación del mensaje electoral llevando a confusión o sobredimensionando la percepción de los electores respecto a la importancia de determinados temas. Finalmente puede generar desinformación, polarización del electorado o propagación de discursos contrarios a los valores democráticos.

Ahora bien, desde el plano regulatorio, cabe destacar que el problema de la manipulación y la desinformación no se agota en la protección de datos[66]. La acertada propuesta regulatoria de la UE pasa por promover medidas dirigidas a fomentar la transparencia y la rendición de cuentas

64 Así, ZUIDERVEEN B., F. (2018), *Ob. Cit.*, p. 85.

65 Ha alertado sobre las nuevas tecnologías y los riesgos de la infra y sobrerrepresentación, COTINO HUESO, L. (2011), «Tratamiento jurídico y normativo de la democracia, participación y transparencia electrónicas: presente y perspectivas», en BARRAT I ESTEVE, J. Y FERNÁNDEZ RIVERA, R., *Derecho de sufragio y participación ciudadana a través de las nuevas tecnologías,* Civitas-Thomson Reuters, Cizur Menor. p. 234-237.

66 Esta es la conclusión del Supervisor Europeo de Protección de Datos. EDPS, *Opinion* 3/2018, p. 13-17. También la de los relatores especiales sobre libertad de expresión de Naciones Unidas y los sistemas regionales de protección de derechos humanos, que han abogado por incorporar un enfoque basado en derechos humanos. Véase la Declaración conjunta sobre libertad de expresión y noticias falsas, desinformación y propaganda de la ONU, OSCE, CIDH y la Comisión Africana de Derechos Humanos, de 3 de marzo de 2017. La misma opinión comparte el Grupo de Alto Nivel de la Comisión Europea sobre *Fake News* y desinformación online en su informe *A multi-dimensional approach to disinformation*, de 12 de marzo de 2018.

en el ecosistema informativo[67], pero sin comprometer la libertad de expresión ni dar pie a injerencias inadmisibles de naturaleza política que afecte la libertad informativa, de expresión o ideológica, o reduzcan la libertad y pluralismo de los medios de comunicación, y, en definitiva, suponga el riesgo de permear la censura de contenidos.

En resumen, el *microtargeting* puede reforzar la movilización electoral y facilitar que los electores tomen decisiones informadas, pero al mismo tiempo genera riesgos de manipulación, desinformación y pérdida de transparencia que pueden afectar la salud del sistema democrático. Aunque nos veamos inclinados a proponer la regulación del perfilado ideológico, bajo la sujeción de garantías como las que comentaremos posteriormente, la valoración última de sus riesgos y ventajas corresponde al Legislador, como depositario de la voluntad popular, que debe aquilatarlo eligiendo además entre un modelo en el que prepondere bien la neutralidad del Estado y la menor interferencia sobre el mercado de ideas, o bien la democracia militante, que no es extraña a la tradición.

La prisa por introducir la regulación mediante la fórmula elegida por el Legislador Orgánico, privó a la figura de un debate saludable con miras a regular, o en su caso proscribir, el perfilado ideológico y conjurar el imparable reto que la vigente y absoluta orfandad regulatoria no puede contener.

VI. Las discriminaciones algorítmicas y su inacabada regulación

Un último problema que pretendemos abordar esta relacionado con la creciente preocupaciones respecto a las decisiones mediadas por algoritmos, que pueden suponer directa o indirectamente un trato discriminatorio. Esta consternación se debe no sólo a la permeabilidad de la mediación algorítmica, sino a que esta expansión progresiva está alcanzando sectores que antes se encontraban bajo el entero control de decisiones humanas, y que tienen un impacto mucho más significativo para la vida y ejercicio de los derechos de los ciudadanos y la sociedad en general[68]. Uno de los argumentos que justifica incorporar algoritmos y *machine learning* es su pretendida neutralidad en comparación con las decisiones humanas que, por diversos motivos, podrían ser sesgadas o prejuiciadas.

67 Comunicación de la Comisión Europea, COM (2018), 236 final, La lucha contra la desinformación en línea, p. 7-17.

68 Por señalar un par de ejemplos, piénsese en el sector financiero, donde los algoritmos de perfilación de usuarios han tenido un impacto importante en la concesión o denegación de hipotecas o préstamos.

Sin embargo, y pese a que a la tecnología se le suele presuponer objetiva y neutra[69], la relación entre tecnología y discriminación no resulta extraña. La introducción de sesgos en el desarrollo tecnológico puede ser resultado de decisiones aparentemente inofensivas pero deliberadas que reflejen prejuicios existentes en la sociedad, o pueden ser consecuencia de errores en el diseño, desarrollo o implementación de la tecnología.

Un ejemplo de lo primero es el desarrollo de la tecnología fotográfica. En su trabajo sobre la representación racial en los medios visuales, Dyer[70] aborda la normalización de una raza, la caucásica, como parámetro de adecuación y óptimo funcionamiento de las cámaras fílmicas y fotográficas. La tecnología en este caso, no es aséptica, en el sentido de la representación racial, puesto que se ha desarrollado, por diversas razones, previendo como usuarios promedios personas de raza blanca. Lo que se aparta de ese canon se considera anormal o problemático[71].

La justificación detrás de esta convención respondió a límites técnicos, pero en definitiva y sobre todo, a una elección que ha generalizado la creencia de que los aparatos fotográficos funcionan mejor con su público normalizado[72].

Sandvig[73], recoge el problema que en el año 2009 tuvo que enfrentar HP con su algoritmo de reconocimiento facial. Concretamente, las cámaras MediaSmart Webcam y el sistema de seguimiento facial incorporado en los ordenadores HP, bajo determinadas condiciones, no reconocían los rostros de personas de color. Este fallo técnico, del que se hicieron eco los medios[74], se hizo viral al publicarse un video en YouTube en el que

69 En la doctrina española se ha ocupado de este tema, entre otros, ESTEVE PARDO, J. (2009), *El desconcierto del Leviatán. Política y Derecho antes las incertidumbres de la Ciencia*, Marcial Pons, Madrid.

70 DYER, R., (1997), *White: Essays on Race and Culture*, Routledge, Londres.

71 El desarrollo tecnológico y las inversiones que le acompañaron, tuvieron como premisa que la población afro-americana no tenía suficiente valor desde el punto de vista fotográfico, a pesar de que la representación de la piel blanca, técnicamente, resultaba más problemática por la tendencia a traducirse en tonos rojizos muy irreales. Sin embargo, una vez superado este escollo técnico, la tez blanca se convirtió en la norma y fotografiar a personas no caucásicas, se consideró problemático o la excepción a la regla. SANDVIG, C. *et al.*, (2016), «Automation, Algorithms, and Politics. When the Algorithm Itself is a Racist: Diagnosing Ethical Harm in the Basic Components of Software», *International Journal of Communication*, 10(0), p. 4973.

72 Como señala DYER, *Ob.* Cit, p. 84, es evidentemente cierto que las cámaras fotográficas representan mejor a las personas blancas. Pero esto es así porque han sido diseñados de esta manera. Así ha sido desde los primeros experimentos realizados con la química de la fotografía, hasta la adopción de las denominadas «Shirley Cards» en la que se representa una modelo caucásica como estándar internacional, promovido por Kodak, para calibrar los ajustes de color e iluminación. Sobre esto véase ROTH, L., (2009), «Looking at Shirley, the Ultimate Norm: Colour Balance, Image Technologies, and Cognitive Equity», *Canadian Journal of Communication*, 34, pp.111–136.

73 SANDVIG, C., *et al.*, *Ob. Cit.*, p. 4973.

74 Véase, por todos, la nota publicado en The Guardian: «Are Hewlett-Packard computer really racist?», disponible en: https://goo.gl/gBYAus; y en CNN: «HP looking into claim webcams can't see black people» https://goo.gl/7w1E49.

se registraba como la cámara seguía sin mayor problema a una persona blanca, mientras que se detenía ante la presencia de una persona de color[75]. Algo similar ocurrió un año después con las Cámaras Nikon, cuyo algoritmo trataba como defectuosas las fotos de asiáticos sonriendo[76].

En controversias similares se ha visto envuelta Google, en relación con algunos de sus algoritmos. En 2015, un algoritmo de inteligencia artificial para el reconocimiento y etiquetado automático de fotos, etiquetó como gorilas a dos personas afroamericanas que se habían realizado un *selfie*[77], lo que obligó a la empresa a pedir disculpas públicamente[78]. También, en 2016, un ciudadano estadounidense, Kabir Alli, puso en evidencia que el buscador de Google arrojaba resultados que, al menos, podría considerarse rozaban el racismo[79]. En concreto, el usuario pedía al buscador que identificara fotos de tres adolescentes negros y tres adolescentes blancos, dando como resultado en el primer caso imágenes de la ficha policial de tres adolescentes detenidos por la policía, mientras que en el segundo presentaba imágenes de tres jóvenes riendo o con equipación deportiva[80].

Las ciencias de la computación han destacado, desde hace ya décadas, que los sistemas computacionales, si se dan determinados condicionamientos, pueden no ser neutrales. Friedman y Nissenbaum, en su estudio pionero ya adelantaron un marco para comprender y sistematizar estas actuaciones prejuiciadas o discriminatorias de los sistemas computacionales[81].

No obstante, en términos generales, la neutralidad que se suele reconocer a la investigación científica se ha transferido al funcionamiento de los algoritmos. Los algoritmos, el *big data* y el *machine learning*, se sostienen sobre una sólida base matemática que no se ve afectada, e. g, por el color de la piel o genero de las personas. A esto hay que sumar el

75 El video, titulado HP computers are racist, sigue disponible en YouTube: https://www.youtube.com/watch?v=t4DT3tQqgRM.

76 Véase el reportaje publicado por Time: «Are Face-Detection Cameras Racist?», disponible en: https://goo.gl/nHdWib.

77 La usuaria Jacky Alcine, hizo la denuncia en Twitter. Puede verse en el siguiente enlace: https://twitter.com/jackyalcine/status/615329515909156865/

78 Véase la nota de la BBC: «Google apologises for Photos app's racist blunder», http://www.bbc.com/news/technology-33347866; o The Guardian: «Google says sorry for racist auto-tag in photo app», https://www.theguardian.com/technology/2015/jul/01/google-sorry-racist-auto-tag-photo-app.

79 https://twitter.com/Q_Piece/status/714821429439561728/video/1.

80 La noticia fue recogida también por los medios. Véase la nota publicada por The Guardian: https://goo.gl/nXyqUK.

81 FRIEDMAN, B. y NISSENBAUM, H. (1996), «Bias in Computer Systems», *ACM Transactions on Information Systems*, Vol. 14, Núm. 3, p. 330 y ss.

sesgo por la automatización y la complacencia tecnológica, que lleva a las personas a confiar acríticamente en las decisiones automatizadas, incluso en supuestos en que sospecha que puede haber un fallo o error en el funcionamiento del sistema[82].

Algunos autores, no obstante, han demostrado que no es difícil desmontar esta presunción de neutralidad que se amplifica por el sesgo por la automatización. Per se, ni los algoritmos ni recurrir al *big data* y el *machine learning* resultan neutrales por defecto. Por el contrario, pueden representar una nueva fórmula para desarrollar prácticas discriminatorias o potenciar la discriminación institucionalizada. Sin embargo, la complejidad técnica ha servido de escudo para que durante mucho tiempo el debate sobre estas prácticas se mantuviera soterrado.

Pueden reconducirse a dos los escenarios que pueden dar lugar a discriminación o decisiones sesgadas. En primer lugar, existe un escenario de búsqueda, obtención o utilización de la información con fines de desarrollar prácticas discriminatorias de manera directa o indirecta. En este sentido, especialmente el *big data* y *el machine learning* podrían actuar como una suerte de baipás para evitar la normativa antidiscriminatoria.

Esta normativa se ha venido consolidando, en el sentido que estudiamos, mediante la protección de datos que se consideran sensibles porque pueden utilizarse para identificar grupos tradicionalmente discriminados o, más aún, para una vez identificados usarse para adoptar decisiones discriminatorias o que los coloque en una situación de desventaja.

En este caso, el *big data* y el *machine learning* pueden servir de instrumentos para acceder a esta información sensible protegida, mediante inferencias a partir de datos relacionados. En otras palabras, pueden representar una tecnología facilitadora que habilita para obtener información protegida a partir de inferencias y, en consecuencia, desarrollar prácticas discriminatorias.

El segundo escenario, por el contrario, no busca de manera directa o indirecta nuevas fórmulas de discriminación, mediante la manipulación de algoritmos o recurriendo a las técnicas de *big data* o *machine learning*. En este caso, el algoritmo o la utilización de *big data* o *machine learning*, de forma no intencionada reproduce un perjuicio o una práctica discriminatoria al recoger fallos en los modelos aplicados, porque la información utilizada recoge sesgos estadísticos o explícita diferencias culturales que son explotadas involuntariamente.

Ahora bien, las posibilidades de que estos escenarios se manifiesten, dependen de determinados condicionamientos técnicos, que conviene tener presentes. En este sentido, algunos autores[83] ha señalado que existen precondiciones técnicas que deben ser consideradas si se pretende evitar que las decisiones algorítmicas sean neutrales, tanto en un caso como en otro. Estas precondiciones pueden agruparse en tres supuestos, que mencionaremos seguidamente.

En primer lugar, los algoritmos pueden considerarse espejos de las discriminaciones o prejuicios preexistentes. En este sentido, los sistemas informáticos como los algoritmos pueden ser espejos sociales[84]. Se trata pues de prejuicios preexistentes[85], pues su origen lo encontramos en las actitudes y prácticas sociales e institucionales, o reflejan prejuicios personales de los diseñadores del sistema o de aquellos actores que pueden influir en su diseño de manera determinante. Por tanto, no derivan del funcionamiento del sistema informático o el algoritmo mismo.

Este sería el caso, por ejemplo, de un algoritmo para determinar la capacidad crediticia de solicitantes de hipoteca, y el diseñador del algoritmo de manera automática califique negativamente a las personas que residan en zonas consideradas indeseables, por concentrar altos índices de población con bajos ingresos o altos índices de criminalidad. En este caso, el diseñador está ordenando al algoritmo que refleje su prejuicio excluyendo las solicitudes de hipotecas de los grupos que ha estereotipado previamente.

Otra forma de reflejar las posibles discriminaciones institucionalizadas a nivel social, pueden derivar no ya de los propios prejuicios del diseñador del algoritmo, sino de los datos utilizados. En la medida en que los *datasets* utilizados para entrenar los algoritmos incorporen sesgos discriminatorios o prejuicios hacía determinados grupos minoritarios, el algoritmo necesariamente los incorporará para, e. g., adoptar su decisión o realizar una clasificación.

82 Las conclusiones de los pioneros estudios sobre el sesgo por la automatización en el sector de la aviación recogidos entre otros, en PARASURAMAN, R. & RILEY, V., (1997), «Humans and Automation: Use, Misuse, Disuse, Abuse. Human Factors», en *The Journal of the Human Factors and Ergonomics Society*, 39(2), pp. 239; SKITKA, L.J. *et al.*, (2000), «Automation bias and errors: are crews better than individuals?», *The International Journal of Aviation Psychology*, 10(1), pp.85–97, son perfectamente aplicables a los sectores de desarrollo algorítmico.

83 Así, HARDT, M., «How Big Data is Unfair. Understanding unintended sources of unfairness in data driven decision making», en medium.com, https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de.

84 Esta la explicación que comentaristas y periodistas especializados dieron respecto a los resultados prejuiciados del buscador de Google respecto a los tres adolescentes negros. Así, YORK, C., «Three Black Teenagers: Is Google Racist? It's not them, it's us», The Huffington Post UK, 08/06/2016, disponible en: https://goo.gl/TYzeTR. También ALLEN, A., «The 'three black teenagers' search shows it is society, not Google, that is racist», The Guardian, 10/08/2016, disponible en: https://goo.gl/vgJHpC.

85 Hablamos de prejuicios preexistentes en el sentido señalado por Friedman y Nissenbaum: Se trata de supuestos en los que los sistemas informáticos reflejan prejuicios que existen independientemente y previos a la creación del sistema. Aunque se refiere a sistemas informáticos, esta definición se puede extrapolar perfectamente a los algoritmos. FRIEDMAN, B. y NISSENBAUM, H., *Ob, Cit.*, p. 333.

Un argumento en contra de este posible sesgo es que sí la data en sí misma no contiene ninguna referencia explícita a algún elemento que pueda ser objeto de prejuicio o discriminación, entonces el algoritmo no podría incluirlo en su resultado. No obstante, los avances de la inteligencia artificial, se dirigen precisamente a extraer atributos (como serían, entre otros, raza o sexo) que no se encuentren explícitamente incorporados en los datos. Estos datos, como ha señalado Hardt, se encuentren latentes en la data y nada evitaría que un algoritmo de aprendizaje los descubra[86].

En segundo término, también relacionado con la data, las discriminaciones o sesgos pueden derivar de la disparidad y calidad de la información recogida en la muestra o datos de entrenamiento respecto a grupos minoritarios. En términos simples, por regla general proporcionalmente siempre existe menos información sobre minorías que respecto a los grupos de población mayoritarios[87]. Por tanto, las decisiones o clasificaciones adoptadas por los algoritmos no representan de manera fiel a los primeros.

La consecuencia de esta disparidad es que las decisiones algorítmicas adoptadas bajo estas condiciones tenderán a ser más ajustadas respecto a los grupos estadísticos dominantes, y más errática respecto a los grupos minoritarios. Gráficamente esto supondría, utilizando como ejemplo la clasificación realizada para la admisión en una universidad, el equivalente a lanzar una moneda al aire respecto a minorías, y decisiones concienzudas y equilibradas en caso de grupos mayoritarios.

Finalmente, otro aspecto que puede permitir que las decisiones algorítmicas sean prejuiciadas o no neutrales, radica en la presencia o posible identificación y explotación de diferencias culturales que permitan determinar la pertenencia a un grupo minoritario estereotipándolo.

El uso de nombre distintivos o poco comunes como rasgos de identidad pueden ser un elemento para determinar la pertenencia a grupos étnicos o minoritarios. Por ejemplo, son numerosos los estudios que han determinado que el uso de nombres peculiares o étnicos entre población afroamericana estadounidense ha sido una fórmula para discriminar candidatos en ofertas de empleo[88]. En estos casos, la discriminación puede

86 HARDT, Moritz, *Ob. Cit.*, p. 2.

87 HARDT, Moritz, *Ob. Cit.*, p. 3

88 BERTRAND, M. y MULLAINATHAN, S. (2004), «Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination», *The American Economic Review*, 94(4), pp. 991–1013. Cuestiona la conclusión de que la discriminación se basa en estos casos en razones exclusivamente raciales, BANKS, R.R. (2009), «Class and Culture: The Indeterminacy of Nondiscrimination», *Stanford Journal of Civil Rights & Civil Liberties*, No. 5, p. 11-15.

basarse simplemente en motivos raciales, o como algunos autores han señalado, a partir de estos nombres étnicos pueden extraerse conclusiones sobre la situación socio-económica y el nivel cultural de los individuos que representan el conjunto de causas del trato discriminatorio[89].

A nivel global, y centrado especialmente en decisiones algorítmicas, las políticas de identificación con nombres verdaderos, también conocidas como Nymwars, han demostrado que los algoritmos pueden considerar falsos nombres típicamente étnicos, generando información errónea o prejuiciada sobre estos grupos. La decisión de Facebook, así como de otras redes sociales, de revisar o suspender las cuentas de aquellos usuarios cuyos nombres parecieran falsos son una muestra de que las diferencias culturales permiten identificar, y en su caso someter a un trato diferenciado, a las personas cuyos nombres no son representativos de los grupos mayoritarios[90].

Por tanto, cuestiones etnográficas o rasgos culturales asociados a grupos minoritarios pueden permitir identificar, incluso involuntariamente a ciudadanos pertenecientes a grupos minoritarios, aunque los algoritmos no hayan sido programados para discriminar, o afectar derechos fundamentales.

Ahora bien, pese a este panorama la respuesta jurídica a esta problemática aún se encuentra en una fase excesivamente germinal, por lo que el análisis técnico-jurídico resulta imposible de abordar. Es cierto que instituciones como el Consejo de Europa[91] o la Unión Europea han expresado su preocupación, e incluso iniciado procesos de creación de grupos de expertos o comisiones alto nivel que abanderen propuestas, pero a la fecha nos encontramos huérfanos de criterios ciertos y resultados tangibles para enfrentarnos a esta tarea.

La experiencia regulatoria más avanzada, se está desarrollando en Estados Unidos con la aprobación de la *Algorithmic Accountability Act* de

89 Como señala BANKS, R.R., *Ob. Cit.*, p.1, las prácticas discriminatorias no suelen basarse exclusivamente en la raza. Se asocian a factores socioculturales, lo que dificulta, en no pocas ocasiones, determinar si estamos ante una situación discriminatoria.

90 Tal fue el caso de las suspensiones de Facebook de nombres distintivos nativoamericanos. Así ocurrió con Lance Browneyes, ciudadano nativo americano cuya cuenta de Facebook fue considerada falsa por los algoritmos de esta red social, y habilitada posteriormente una vez comprobada su identidad, no sin antes modificar su apellido a uno más habitual entre la población estadounidense: Brown. Casos similares se presentaron con otros ciudadanos, lo que llevó a los medios a considerar que Facebook, con carácter general, estaba cuestionando la legitimidad de los nombres nativo americanos. Véase sobre este caso la nota publicada por la BBC: https://goo.gl/nGLsVz. Notable fue también el caso del autor Salman Rushdie, cuya cuenta fue suspendida y su nombre modificado por el de Ahmed, recogido también por medios internacionales dando lugar a una rápida modificación por Facebook. Véase, Rushdie Runs Afoul of Web's Real-Name Police, New York Times, 14 de noviembre de 2011. Disponible en: http://www.nytimes.com/2011/11/15/technology/hiding-or-using-your-name-online-and-who-decides.html.

91 Véase: https://www.coe.int/en/web/artificial-intelligence/cahai

2019, que posiblemente servirá de inspiración a efectos de la discusión que se desarrollará en el ámbito europeo, pero que excede jurisdiccionalmente los objetivos de nuestra investigación.

Sirva en todo caso este apartado para dejar constancia de la urgencia de abordar y conjurar mediante desarrollo normativo los riesgos que pueden derivar de decisiones algorítmicas discriminatorias.

VII. Bibliografía

Banks, R.R. (2009), «Class and Culture: The Indeterminacy of Nondiscrimination», *Stanford Journal of Civil Rights & Civil Liberties*, Núm. 5.

Bertrand, M. y Mullainathan, S. (2004), «Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination», *The American Economic Review*, 94(4).

Bosco, F., *et. al* (2015), «Profiling Technologies and Fundamental Rights. An Introduction», en Creemers, N., et.al (ed.), *Profiling Technologies in Practice. Applications and Impact on Fundamental Rights and Values*, WLP, Oisterwijt.

Buczak, A. L. y Guven, E. (2016) «A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection», *IEEE Communications Surveys and Tutorials*, IEEE, 18(2).

Caballero, R. y Martín, E. (2015), *Las bases del Big Data,* Catarata, Madrid

Clarke, R. (1993), «Profiling: A Hidden Challenge to the Regulation of Data Surveillance», *Journal of Law, Information and Science,* Volumen 4, Núm. 2.

Coglianese, C. & LEHR, D. (2017), *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*.

Cotino Hueso, L. (2011), «Tratamiento jurídico y normativo de la democracia, participación y transparencia electrónicas: presente y perspectivas», en BARRAT I ESTEVE, J. Y FERNÁNDEZ RIVERA, R., *Derecho de sufragio y participación ciudadana a través de las nuevas tecnologías,* Civitas-Thomson Reuters, Cizur Menor.

Dada, E. G., *et al*. (2019) «Machine learning for email spam filtering: review, approaches and open research problems», en *Heliyon*, Elsevier, 5(6).

Deng, L. and LI, X. (2013) «Machine learning paradigms for speech recognition: An overview», *IEEE Transactions on Audio, Speech and Language Processing*, 21(5).

Dyer, R., (1997), *White: Essays on Race and Culture*, Routledge, Londres.

Edwards, L. y Veale, M. (2018), «Enslaving the algorithm: from a 'right to an explanation' to a 'right to better decisions'? », *IEEE Security & Privacy*, Vol. 16, Núm. 3.

Esteve Pardo, J. (2009), *El desconcierto del Leviatán. Política y Derecho antes las incertidumbres de la Ciencia*, Marcial Pons, Madrid.

Friedman, B. y NISSENBAUM, H. (1996), «Bias in Computer Systems», *ACM Transactions on Information Systems*, Vol. 14, Núm. 3.

García-Pelayo, M. (1986), *El Estado de Partidos*, Alianza, Madrid.

Goodman, B. y Flaxman, S. (2017), «European Union regulations on algorithmic Decision-making and a right to explanation», *AI Magazine*, Vol. 38, Núm. 3.

Guillory, J.E. et al., (2014), «Experimental evidence of massive scale emotional contagion through social networks», *Proceedings of the National Academy of Sciences*, 111(29).

Hernández López, J.M. (2013), *El Derecho a la protección de Datos Personales en la Doctrina del Tribunal Constitucional*, Thomson Reuters-Aranzadi, Cizur Menor.

Hildebrant, M. (2008), «Defining Profiling: A New Typo of Knowledge?», en HILDEBRANT M., y Gutwirth, S., *Profiling the European Citizen: Cross-Disciplinary Perspectives*, Springer, Dordrecht.

Hoffman-Riem, W., (2004), «Methoden einer anwendungsorientierten Verwaltungsrechtswissenschaft», en SCHMIDT-ASSMANN, E. y HOFFMAN-RIEM, W. (ed.), *Methoden der* Verwaltungsrechtswissenschaft, Nomos, Baden Baden.

Kelleher, J.D., *et. al* (2015), *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies,* The MIT Press, Cambridge.

Laney, D. (2001), «3-D Data Management: Controlling Data Volume, Velocity and Variety», *META Group Research Note*.

Lucas Murillo, P. (1990), *El derecho a la autodeterminación informativa*, Tecnos, Madrid.

Mantelero, A. (2017), «Regulating big data. The guidelines of the Council of Europe in the context of the European data protection framework», en *Computer Law and Security Review*, 33(5).

Martínez Cuevas, M. D. (2006), *Régimen jurídico de los partidos políticos*, Marcial Pons, Madrid.

Mitchell, T. (2006), *The Discipline of Machine Learning*, Carnegie Mellon University, CMU-ML-06-108.

Munoz, C., *et. al*, (2016), «Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights», *Executive Office of the President of USA*.

Neapolitan, R.E y JIANG, X. (2018), *Artificial Intelligence. With an Introduction to Machine Learning*, segunda edición, CRC Press, Boca Ratón

Palma Ortigosa, A. (2019), «Decisiones automatizadas en el RGPD. El uso de algoritmos en el contexto de la protección de datos», en *Revista General de Derecho Administrativo*, Núm. 50.

Parasuraman, R. & Riley, V., (1997), «Humans and Automation: Use, Misuse, Disuse, Abuse. Human Factors», *The Journal of the Human Factors and Ergonomics Society*, 39(2).

Pasquale, F. (2015), *The Black Box Society. The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge.

Podesta, J. *et al.*, (2014), «Big Data: Seizing Opportunities», *Executive Office of the President of USA*.

Polakiewicz, J. (2011), «Convention 108 as a global privacy standard?», en *International Data Protection Conference*, Budapest.

Portugal, I., *et al*. (2018), «The use of machine learning algorithms in recommender systems: A systematic review», *Expert Systems with Applications*, Elsevier, 97.

Presno Linera, M. A. (2000), *Los partidos políticos en el sistema constitucional español. Prontuario de jurisprudencia constitucional: 1980-1999*, Aranzadi, Pamplona.
Roth, L., (2009), «Looking at Shirley, the Ultimate Norm: Colour Balance, Image Technologies, and Cognitive Equity», *Canadian Journal of Communication*, Núm. 34.
Ruiz Miguel, C. (2003), «El derecho a la protección de los datos personales en la Carta de Derechos Fundamentales de la Unión Europea: Análisis crítico», *Revista de Derecho Comunitario Europeo*, Año 7, Núm. 14, enero-abril.
Russell, S. Y Norvig, P. (2016), *Artificial Intelligence. A Modern Approach,* tercera edición, Pearson, Essex.
Samuel, Arthur L. (1959), «Some Studies in Machine Learning Using the Game of Checkers», *IBM Journal*, Vol. 3, Núm. 3.
Sandvig, C. *et al.*, (2016), «Automation, Algorithms, and Politics. When the Algorithm Itself is a Racist: Diagnosing Ethical Harm in the Basic Components of Software», *International Journal of Communication*, 10(0).
Serrano Pérez, M.M. (2003), *El derecho fundamental a la protección de datos. Derecho español y* comparado, Thomson-Civitas, Cizur Menor.
Skitka, L.J. *et al.*, (2000), «Automation bias and errors: are crews better than individuals?», *The International Journal of Aviation Psychology*, 10(1).
Stilgoe, J. (2018), «Machine learning, social learning and the governance of self-driving cars», *Social Studies of Science*, 48(1).
Taylor, L., *et.al*, (eds.), (2017), *Group Privacy New Challenges of Data Technologies*, Springer, Dordrecht.
Virgala Foruria, E. (2017), *La regulación jurídica de los partidos políticos*, Ediciones Olejnik, Santiago de Chile.
Watcher. S., Mittelstadt, B. y FLORIDI, L. (2017), «Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation», *International Data Privacy Law*, Vol. 7, Núm. 2.
Zuiderveen B., F. J. (2018), «Online Political Microtargeting: Promises and Threats for Democracy», *Utrecht Law Review*, Vol. 14, Núm. 1.

Encarte