Revista Española de la Transparencia
RET. ISSN 2444-2607
Núm. 17. Número Extraordinario 2023
www.revistatransparencia.com

Sección Estudios y Artículos
Las exigencias de transparencia para los sistemas algorítmicos…
por Ana Garriga Domínguez
Págs. 137-164
DOI: https://doi.org/10.51915/ret.309

Las exigencias de transparencia para los sistemas algorítmicos de recomendación, selección de contenidos y publicidad en línea en el nuevo Reglamento Europeo de Servicios Digitales

Ana Garriga Domínguez[1]

Universidad de Vigo
España

ORCID: 0000-0003-2846-3448

RECIBIDO : 28/04/2023
ACEPTADO : 7/07/2023

Este trabajo se ha realizado en el marco del proyecto de investigación DATATRANSCO con referencia n.º PID2021-128309NB-I00.

RESUMEN: En este trabajo se recogen las principales problemáticas que suscitan los sistemas de inteligencia artificial, que determinados prestadores de servicios en Internet emplean para personalizar la publicidad en línea y para seleccionar y recomendar información y contenidos. Se estudian las obligaciones derivadas del Reglamento General de Protección de Datos, con particular atención al principio de responsabilidad proactiva y transparencia, en relación con esos sistemas. Finalmente, se abordan las novedades que, en materia de transparencia, introduce el nuevo Reglamento Europeo de Servicios Digitales. Particularmente se estudian las nuevas obligaciones de transparencia algorítmica y rendición de cuentas para los motores de búsqueda en línea de muy gran tamaño y las plataformas en línea de muy gran tamaño, considerando que el nuevo Reglamento ofrece herramientas muy útiles y garantías importantes para luchar contra la desinformación y proteger el derecho a recibir informaciones plurales, la libertad de expresión e ideológica, así como para prevenir la difusión de contenido ilícito o dañino.

[1] Profesora Titular de Filosofía del Derecho y, desde 2019, Delegada de Protección de Datos de la Universidad de Vigo. Su labor docente mayoritaria la realiza en la Escuela Superior de Ingeniería Informática de esta Universidad, de la que ha sido Directora. Es, asimismo, Directora del Laboratorio "Sociedad de la Información y derechos humanos" de la Universidad de Vigo (Laboratorio Consolider-Ingenio 2010 "El tiempo de los derechos"). Ha sido Investigadora principal en numerosos Proyectos de investigación y colaboradora con instituciones públicas y privadas en actividades de I+D y es autora de numerosas publicaciones sobre derechos fundamentales, sociedad de la información y protección de datos personales.

Revista Española de la Transparencia
RET. ISSN 2444-2607
Núm. 17. Número Extraordinario 2023
www.revistatransparencia.com

Sección Estudios y Artículos
Las exigencias de transparencia para los sistemas algorítmicos…
por Ana Garriga Domínguez
Págs. 137-164
DOI: https://doi.org/10.51915/ret.309

PALABRAS CLAVE: Inteligencia artificial, sistemas de recomendación, transparencia algorítmica, protección de datos personales, desinformación.

CONTENIDOS: 1.-Sistemas algorítmicos de recomendación, selección de contenidos y publicidad en línea. Riesgos para los derechos fundamentales. 2.-Las garantías derivadas del principio de transparencia en el Reglamento General de Protección de Datos. *2.1.-Decisiones automatizadas y elaboración de perfiles. 2.2.-La importancia de los principios de transparencia y de responsabilidad proactiva del RGPD*[2] *en los sistemas de IA.* 3.-Obligaciones específicas para VLOP y VLOSE en el Reglamento (UE) 2022/2065, de 19 de octubre, de Servicios Digitales (RSD). Especial referencia a las obligaciones de transparencia algorítmica. *3.1.- Un Reglamento para crear un entorno en línea seguro y frenar la difusión de desinformación y manipulación en línea. 3.2.- Obligaciones de transparencia para los motores de búsqueda (VLOSE) y las plataformas en línea de muy gran tamaño (VLOP). Especial referencia a las obligaciones de transparencia algorítmica.* 4.-Conclusión.

Transparency requirements for algorithmic recommender systems, content selection and online advertising in the new European Digital Services Act

ABSTRACT: This paper addresses the main issues raised by artificial intelligence systems, which are used by certain Internet service providers to personalise online advertising and to select and recommend information and content. It examines the obligations arising from the General Data Protection Regulation (GDPR), with particular attention to the principle of proactive liability and transparency, in relation to these systems. Finally, it deals with the new developments introduced by the new European Digital Services Act (DSA) in the field of transparency. In particular, the new algorithmic transparency and accountability obligations for very large online search engines and very large online platforms are studied, considering that the new Regulation offers useful tools and important guarantees to fight disinformation and protect the right to receive pluralistic information, freedom of expression and ideology, as well as to prevent the dissemination of illegal or harmful content.

KEYWORDS: Artificial intelligence, recommender systems, algorithmic transparency, personal data protection, disinformation.

[2] REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos).

Revista Española de la Transparencia
RET. ISSN 2444-2607
Núm. 17. Número Extraordinario 2023
www.revistatransparencia.com

Págs. 137-164
DOI: https://doi.org/10.51915/ret.309

1.- Sistemas algorítmicos de recomendación, selección de contenidos y publicidad en línea. Riesgos para los derechos fundamentales

En los últimos años se ha producido un espectacular desarrollo tecnológico, que ha tenido como protagonista el avance en los sistemas de Inteligencia Artificial (IA). Su desarrollo es comparable al que, en las dos décadas anteriores, experimentó Internet y con el vertiginoso incremento del número y clase de servicios disponibles a través de la red usados por millones de personas. Su evolución, desde la IA simbólica, que se basaba en reglas predefinidas ejecutadas por una máquina, a la IA que se sirve de grandes conjuntos de datos (Big Data) y que se basa en el aprendizaje automático, ha tenido como consecuencia el incremento de su número de aplicaciones y de su capacidad para resolver problemas. Pero, su desarrollo y aplicación no ha estado exenta de críticas y cuestionamientos por los problemas que muchos de estos sistemas plantean por su injerencia en los derechos fundamentales y los valores democráticos. Un ejemplo muy reciente lo tenemos en el sistema de IA generativa, *ChatGPT*[3]*,* que permite a cualquier usuario mantener conversaciones, producir textos sobre temáticas diversas, facilitar respuestas planteadas por los usuarios (en ocasiones exactas, en otras con errores e inexactitudes) y que tiene un sentido del contexto que le permite mantener una conversación con su interlocutor coherente con su interacción con ella. Se trata de un sistema que puede realizar multitud de tareas, desde escribir líneas de código a redacciones o poemas y, por supuesto, mantener una conversación sobre cualquier temática. Desde su lanzamiento en abierto, hemos pasado del asombro y la expectación a la preocupación por sus efectos, en el ámbito educativo, respecto de los derechos de propiedad intelectual, por los riesgos éticos en muchos ámbitos y por su capacidad para difundir desinformación, hasta el punto de que más de un millar de expertos en IA han pedido una moratoria de seis meses[4] para "la carrera sin control de los *ChatGPT*"[5]. Y si bien, esa petición se ha cuestionado por determinados séctores[6], lo

[3] Según se describe por su desarrollador, OpenAI, ChatGPT es un modelo lingüístico diseñado para responder a consultas basadas en texto y generar respuestas en lenguaje natural. Forma parte del campo más amplio de la inteligencia artificial conocido como procesamiento del lenguaje natural (PLN), que trata de enseñar a los ordenadores a entender e interpretar el lenguaje humano.

ChatGPT se basa en una arquitectura de aprendizaje profundo denominada Transformer, que le permite aprender patrones lingüísticos y generar textos coherentes y similares a los humanos. Puede consultarse en su web: GPT Chatbot: Advanced AI Chat.

[4] El texto de la *carta* puede consultarse en: https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

[5] Véase el reportaje publicado en el diario El País (elpais.com) el 29 de marzo de 2023 bajo el título "Expertos en inteligencia artificial reclaman frenar seis meses la "carrera sin control" de los ChatGPTid.

[6] Véase reportaje publicado en el diario La Opinión (laopinioncoruna.es) "Así responden los expertos de A Coruña a la carta de Elon Musk sobre Inteligencia Artificial y ChatGPT" publicado el 9 de abril de 2023.

Revista Española de la Transparencia
RET. ISSN 2444-2607
Núm. 17. Número Extraordinario 2023
www.revistatransparencia.com

Págs. 137-164
DOI: https://doi.org/10.51915/ret.309

cierto es que el pasado 30 de marzo, el *Garante per la Proteziones dei Dati Personali* ordenó la medida de limitación provisional del tratamiento de los datos personales de los interesados establecidos en territorio italiano, al entender que se podían estar produciendo varias infracciones graves tipificadas en el RGPD. También la Agencia Española de Protección de Datos (AEPD) ha iniciado una investigación de oficio a OpenAI, propietaria de ChatGPT[7].

Habida cuenta de la existencia de diferentes definiciones de IA, en este trabajo se proponen dos complementarias. Una más amplia, que es la recogida en la Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO de noviembre de 2021, según la cual, "los sistemas de IA son tecnologías de procesamiento de la información que integran modelos y algoritmos que producen una capacidad para aprender y realizar tareas cognitivas, dando lugar a resultados como la predicción y la adopción de decisiones en entornos materiales y virtuales"; y, otra más estricta y más alineada con la propuesta de la OCEDE, que es la recogida en la última versión de la Propuesta de Reglamento sobre IA[8]. Ésta última, define un sistema de inteligencia artificial como aquel sistema "concebido para funcionar con elementos de autonomía que, a partir de datos e información generados por máquinas o por seres humanos, infiere la manera de alcanzar una serie de objetivos, utilizando para ello estrategias de aprendizaje automático o estrategias basadas en la lógica y el conocimiento, y produce información de salida generada por el sistema, como contenidos (sistemas de inteligencia artificial generativa), predicciones, recomendaciones o decisiones, que influyen en los entornos con los que interactúa el sistema de IA".

La IA tienen múltiples aplicaciones y en muchas de ellas se emplean datos personales en alguna de las etapas de su ciclo de desarrollo y comercialización, ya sea en la fase de entrenamiento y de validación, ya sea en otras etapas posteriores de explotación. Así, por ejemplo, en el ámbito de la salud, de la seguridad, del cálculo del riesgo financiero, etc. No obstante, existen otros muchos sistemas de IA que no los utilizan en ninguna de sus fases de vida, ni en su desarrollo, ni en su explotación, como pueden ser los que se emplean en el ámbito de la geología, la predicción de la climatología, la resistencia de materiales, el diseño industrial, etc. Aunque los requisitos de explicabilidad y transparencia debieran de ser exigibles en cualquier

[7] El anuncio se ha realizado a través de una nota de prensa el día 13 de abril de 2023 y puede consultarse en (aepd.es).

[8] Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión).

DOI: https://doi.org/10.51915/ret.309

sistema de IA[9], en este trabajo se analizarán aquellos que afectan a las personas a través de decisiones automatizadas y que han sido entrenados con datos personales o que los utilizan en alguna de las fases de su ciclo de vida, en un contexto muy concreto, el de las plataformas en línea y motores de búsqueda en línea de muy gran tamaño en el sentido del Reglamento de Servicios Digitales (RSD). Se trata de sistemas de IA que analizan patrones de comportamiento para diseñar publicidad personalizada, ofrecer o seleccionar determinados contenidos o información.

Ha de hacerse referencia en esta introducción a otro factor, que ha condicionado en los últimos años el desarrollo de los sistemas y herramientas del campo de la inteligencia artificial: la cantidad de información personal disponible. Como ya he señalado en otro lugar (Garriga Domínguez, 2015), este hecho nos sitúa por sí mismo ante una nueva revolución tecnológica, el *Big Data,* que "no se cifra en las máquinas que calculan los datos, sino en los datos mismos y en cómo los usamos" (MAYER-SCHÖNBERGER, y CUKIER, 2013: 18). Los avances en la minería y análisis de datos y el aumento masivo de la capacidad informática de procesamiento y almacenamiento han ampliado exponencialmente la información que se encuentra al alcance de terceros, ya sean operadores públicos o privados. Asimismo, el número creciente de personas, dispositivos y sensores que están conectados por redes digitales ha revolucionado la capacidad de generar, comunicar, compartir y acceder a los datos (TENE, y POLONETSKY, 2012: 63). La definición de *Big Data* incluye ambos elementos: por un lado, la gran cantidad de datos disponibles y, por otro, el conjunto de tecnologías cuyo objetivo es tratar esas grandes cantidades de información (BELTRÁN PARDO y SEVILLANO JAÉN. 2013: 16), empleando complejos algoritmos y estadística con la finalidad de hacer predicciones, extraer información oculta o correlaciones imprevistas y, en último término, favorecer la toma de decisiones. El conjunto de tecnologías del ámbito de la IA que se emplean para analizar estas inmensas cantidades de datos, recibe el nombre de «minería de datos» (RAMOS BERNAL, 2012: 186).

Una de las características de muchos de los nuevos sistemas de IA es que utilizan el aprendizaje automático. De acuerdo con la conocida definición de Arthur L. Samuel (SAMUEL, 1983), el proceso de aprendizaje automático de los algoritmos haría referencia *al campo de estudio que da a los computadores la habilidad de aprender algo para lo que no han sido programados expresamente*. Para su desarrollo se necesitan de grandes cantidades de datos de alta calidad. De hecho, muchos de los problemas de sesgo en los sistemas de IA (Vid. O'NEIL, 2017) se deben a datos de entrenamiento de baja calidad, bien porque no son suficientemente diversos, como

[9] Ambos principios se encuentran recogidos en las Directrices éticas para una IA fiable del Grupo independiente de expertos de alto nivel sobre inteligencia artificial, creado por la Comisión Europea en junio de 2018. Asimismo, la Propuesta de Reglamento de Inteligencia Artificial refuerza las obligaciones de transparencia.

Revista Española de la Transparencia
RET. ISSN 2444-2607
Núm. 17. Número Extraordinario 2023
www.revistatransparencia.com

DOI: https://doi.org/10.51915/ret.309

cuando no se incluyen datos de mujeres o de minorías étnicas (Vid. AÑÓN ROIG, 2022), o porque no se cuenta con un volumen suficiente de datos de entrenamiento. Aunque en ocasiones son los propios programadores o desarrolladores quienes introducen sus propios sesgos o prejuicios, en otras se producen fallos porque no se definen correctamente los objetivos del sistema de IA o porque quien debe interpretar los resultados lo hace también de forma sesgada o incorrecta. Precisamente, el tamaño de los datos disponibles en la web, si bien ha permitido a los modelos de aprendizaje profundo alcanzar una alta precisión en aplicaciones PNL, dan lugar a modelos que codifican prejuicios estereotipados y peyorativos en función del género, la raza, la etnia y la discapacidad (BENDER, GEBRU, MCMILLAN-MAJOR y SHMITCHELL, 2021).

Además de los sesgos, que perpetúan los prejuicios y la discriminación (BELLOSO MARTÍN, 2022: 61)[10], otro riesgo de las tecnologías de IA que utilizan *Big Data*, proviene de la posibilidad de extraer patrones de comportamiento y perfiles personales (CRAIG, y LUDLOFF, 2011: 6), haciendo posible una peligrosa y nueva filosofía de la anticipación, cuyo extremo sería el de las predicciones preventivas (KERR y EARLE, 2013: 65). Pueden influir en los deseos de nuevas maneras, pero también puede mediar en los comportamientos reales de ciertos grupos sociales que, como los individuos, son alentados por retroalimentación para ajustarse a los patrones esperados (LYON, 2014: 101). Y, como ha señalado Soshana Zuboff, el matrimonio formado por la modificación de la conducta y los medios tecnológicos adecuados para automatizar su aplicación es fundamental "para la creación de economías de acción" características de estos operadores (ZUBOFF, 2019: 400). Por otra parte, la elaboración de perfiles y el establecimiento de correlaciones y predicciones permitirá visibilizar "modelos colectivos de comportamiento" (BYUNG-CHUL, 2014: 109) y posibilitará la clasificación social de los individuos y los grupos para la adopción de determinadas decisiones sobre ellos.

En este trabajo se analizará la importancia del principio de transparencia y de las exigencias derivadas del derecho a la protección de datos personales respecto de los sistemas de IA utilizados para determinar la información a la que un usuario concreto de una plataforma va a tener acceso. Un ejemplo, lo encontramos en los servicios que utilizan algoritmos para personalizar las noticias u otros contenidos para cada usuario, pues estos emplearán "parámetros de medición basados en afinidades digitales que delimitan, para el usuario, ventanas de visibilidad que tienen el color de su red social" (CARDON, 2018: 43). Así, cada usuario tendrá acceso a una

[10] Señala Nuria Belloso, que "los sesgos algorítmicos se encuentran en todas las plataformas" y que, en la medida en que todo lo que hacemos se procesa y mediatiza por algoritmos sesgados, deben ofrecerse adecuadas respuestas jurídicas a las amenazas a los derechos fundamentales.

Revista Española de la Transparencia
RET. ISSN 2444-2607
Núm. 17. Número Extraordinario 2023
www.revistatransparencia.com

Sección Estudios y Artículos
Las exigencias de transparencia para los sistemas algorítmicos…
por Ana Garriga Domínguez
Págs. 137-164
DOI: https://doi.org/10.51915/ret.309

información determinada en función de sus intereses, de situación particular o de su estado emocional, que habrán sido previamente identificados por el sistema.

Es importante señalar que, el impacto de los sistemas algorítmicos de moderación de contenidos, de recomendación y los sistemas publicitarios personalizados en el ámbito de determinados servicios digitales, como los motores de búsqueda y las plataformas en línea de muy gran tamaño, no se limita al derecho a la privacidad. Los riesgos derivados de determinadas prácticas pueden afectar, asimismo, a las libertades de expresión e información, a la libertad ideológica o los propios valores democráticos. En este contexto y en la medida en que la exposición a noticias, opiniones e información cívica ocurre cada vez más a través de las redes sociales (BAKSHY, MESSING y ADAMIC 2015: 1130) en las que la decisión sobre los contenidos o sobre cómo se presentan se toma en base a perfiles personalizados, los agentes políticos se ven compelidos a utilizar también estos medios por su eficacia y para conseguir una mayor visibilidad (ECHEVERRÍA, 2013: 175 y ss.). Ahora bien, no es lo mismo el debate de ideas o la publicidad legítima (Vid. GARCÍA MAHAMUT, 2015: 30 y ss.), que la propagación de noticias falsas y los fenómenos de desinformación que utilizan el potencial del *Big Data* y de la micro-segmentación para conseguir sus objetivos económicos o políticos[11]. Las posibilidades actuales de micro-segmentación y manipulación *online* basadas en las tecnologías de *Big Data* e inteligencia artificial hacen que el riesgo para los derechos de las personas sea hoy muy elevado.

Como ha señalado el Supervisor Europeo de Protección de Datos (SEPD) existe una amenaza para los valores democráticos y los derechos fundamentales derivados de la incesante vigilancia a la que son sometidas las personas en el espacio digital por empresas y Estados y, esta disminución de su espacio íntimo tiene como consecuencia "un efecto alarmante sobre la capacidad y voluntad de las personas de expresarse y establecer relaciones con libertad, también en la esfera cívica, tan esencial para la salud de la democracia"[12]. Cuando el entorno *online* se encuentra personalizado y micro-segmentado, los ciudadanos estamos expuestos a informaciones que refuerzan los sesgos ideológicos y es más difícil encontrar opiniones diferentes, lo que lleva "a una mayor polarización política e ideológica"[13]. En su Informe de enero de 2018, el Grupo Consultivo sobre Ética del Supervisor Europeo de Protección de Datos señalaba, entre las amenazas para la autonomía

[11] Final report of the *High Level Expert Group on Fake News and Online Disinformation: "A multi-dimensional approach to disinformation"* (12 de marzo de 2018), en: https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation, *p. 11.*

[12] SEPD. *Opinion 3/2018, on online manipulation and personal data*, adoptada el 13 de marzo de 2018, p.3. Este documento está disponible en la página oficial del Supervisor Europeo de Protección de Datos (edps.europa.eu).

[13] *Ibídem,* p. 7.

Revista Española de la Transparencia
RET. ISSN 2444-2607
Núm. 17. Número Extraordinario 2023
www.revistatransparencia.com

Sección Estudios y Artículos
Las exigencias de transparencia para los sistemas algorítmicos…
por Ana Garriga Domínguez
Págs. 137-164
DOI: https://doi.org/10.51915/ret.309

individual, la difusión algorítmica o humana de noticias falsas, que debilita la capacidad de los individuos para discriminar entre lo que es información fiable y lo que no lo es y, así también, los procesos democráticos estarían en riesgo de debilitarse a través de las prácticas de marketing político basadas en técnicas de micro-segmentación o focalización (targeting) y elaboración de perfiles psicográficos[14]; pues, las técnicas de micro-segmentación en el ámbito electoral cambia las reglas del discurso político, reduciendo el espacio para el debate y el intercambio de ideas[15].

La focalización va a permitir transmitir mensajes específicamente diseñados para promover intereses económicos o comerciales, ideológicos o políticos o de cualquier otra clase. Pues, el *targeting* tiene, como objetivo prioritario, orientar o dirigir al sujeto o a un grupo de personas en un sentido y con una finalidad determinados. Como estos sistemas de IA, como la mayoría de los sistemas automatizados de toma o de ayuda a la decisión, resultan tan complejos y opacos para un usuario medio, el principio de transparencia debe tener un papel central a fin de ayudar a comprender por qué se le ofrecen determinados contenidos y no otros.

El Comité Europeo de Protección de Datos (CEPD) ha identificado una larga una serie de riesgos para los derechos y libertades de los usuarios de las plataformas y medios sociales. La elaboración de perfiles socava la capacidad de las personas "para ejercer el control sobre sus datos personales"[16], aumenta el riesgo de discriminación, utilizando, o no, informaciones sensibles, y de manipulación que podría afectar a cuestiones y procesos políticos, al acentuar vulnerabilidades y emociones negativas afectando a la autonomía, la libertad y a la salud psicológica, en especial en el caso de los menores, pues se puede aprovechar determinados momentos en los que el análisis de la información revele determinados estados emocionales "para dirigir a la persona mensajes específicos y en momentos concretos a los que se espera que sea más receptiva e influir así subrepticiamente en su proceso de pensamiento, sus emociones y su comportamiento"[17].

El desarrollo de la IA permite el perfilado ideológico individual, de la misma forma en la que pueden inferirse perfiles de otras clases, ya sea sobre preferencias de

[14] Ethics Advisory Group. 2018. Report 2018, p.18. El informe puede descargarse de la página oficial del Supervisor Europeo de Protección de Datos (dps.europa.eu).

[15] Ibídem, p. 28.

[16] Directrices 8/2020 sobre la focalización de los usuarios de medios sociales, Versión 2.0, adoptadas el 13 de abril de 2021, p. 7.

[17] *Ibídem*, p. 7.

DOI: https://doi.org/10.51915/ret.309

consumo, fiabilidad financiera, emocional[18] o, incluso sobre orientación sexual (SARIGOL, GARCÍA y SCHWEITZER, 2014: 105). La elaboración de perfiles en las plataformas sociales permite aplicar técnicas de micro-segmentación para elaborar información política personalizada. Está clara la relación entre la elaboración de perfiles emocionales e ideológicos y las *fake news* y el fenómeno de la desinformación a través de la red y la institución de la opinión pública libre (ALCARÁZ RAMOS, 2020). En consecuencia, se produce una estrecha conexión entre la garantía de la privacidad de las personas y el ejercicio de estas libertades, que están en la base del sistema democrático.

El escándalo Facebook-Cambridge Analytica evidenció una serie de prácticas que habrían afectado, al menos, a 50 millones de personas y, sobre cuyos datos personales almacenados por Facebook, se habrían elaborado perfiles individuales con fines de focalización política en las elecciones presidenciales de Estados Unidos de 2016 y en el referéndum sobre la permanencia en la Unión Europea del Reino Unido[19]. Cuando la ciudadanía ejerce su derecho al sufragio elige sobre la base de un juicio que se construye sobre el conocimiento del que disponga de los asuntos públicos y su gestión. Y este conocimiento sobre asuntos de relevancia pública puede garantizar esa actuación libre de los ciudadanos pues, como nos recuerda el Tribunal Constitucional, "únicamente aquellas sociedades que pueden recibir informaciones veraces y opiniones diversas de cuanto constituyen los aspectos más importantes de la vida comunitaria, están en condiciones de ejercitar, después, sus derechos y cumplir sus deberes como ciudadanos, partiendo del principio esencial de que la soberanía nacional reside en el pueblo, del que emanan los poderes del Estado"[20]. Pero, la libre circulación de opiniones e informaciones se ve obstaculizada por *bots* y noticias falsas, pero también, cuando se aplican «burbujas de filtro», que a través de filtros invisibles, nos aísla sin percibirlo y el sistema se aprovecha del incentivo psicológico que supone el sesgo de confirmación (ALCOTT, y GENTZKOW, 2017: 211). Al desconocer la forma y los criterios según los cuales los servicios filtran la información que entra y sale, "es prácticamente imposible ver lo sesgada que es" (PARISIER, 2017: 18). Como consecuencia de estas prácticas, también la libertad ideológica podrá resultar afectada. Su papel es esencial en un Estado democrático y presupuesto del derecho de participación, del pluralismo político (XIOL RÍOS, 2001: 18 y ss.) y "requisito de funcionamiento del Estado democrático" (ROLLNERT LIERN, 2002: 70). Para garantizar este elemento negativo de la libertad ideológica y de

[18] A través de las informaciones que los usuarios suben a las redes sociales es posible hacer gráficas de sus emociones, sentimientos y estados de ánimo. Por ejemplo, Twitter "permite la datificación de pensamientos, estados de ánimo e interacciones de la gente" (MAYER-SCHÖNBERGER y CUKIER. 2013: 116 - 117.

[19] De forma detallada se recogen en el Informe de la Cámara de los Comunes de 14 de febrero de 2019 *Disinformation and 'fake news': Final Report".*

[20] STC 173/1995, de 21 de noviembre.

DOI: https://doi.org/10.51915/ret.309

conciencia[21], tanto el RGPD como la LOPDGDD[22] prohíben, como regla general, aunque con determinadas excepciones, el tratamiento de los datos personales que revelen las opiniones políticas[23], las convicciones religiosas o filosóficas (artículos 9.1 de ambas normas). Ahora bien en el ámbito del *Big Data*, las posibilidades de elaboración de perfiles con el auxilio de la inteligencia artificial hace posible inferir las convicciones ideológicas y de conciencia de una persona sin que esta las haya hecho públicas, pudiendo "hallarse correlaciones que indiquen algo sobre la salud, las convicciones políticas, las creencias religiosas o la orientación sexual de las personas"[24].

Puede constatarse que los instrumentos del RGPD son insuficientes ante esta problemática y, en tanto no se apruebe definitivamente el Reglamento de IA, el Reglamento de Servicios Digitales debería servir para afrontar y mitigar los riesgos y problemática descritos. Sus normas suponen un refuerzo a las garantías que existían hasta ahora en el ordenamiento jurídico europeo para combatir éstos y otros problemas que ocasionan los sistemas de publicidad *online* y los sistemas de recomendación y selección de contenidos de los grandes operadores en la web. En este trabajo se prestará atención a las que buscan garantizar una mayor transparencia de los algoritmos usados con esos fines; si bien, como no es necesario explicar, su contenido es mucho más amplio y regula otras muchas cuestiones.

2.- Las garantías derivadas del principio de transparencia en el reglamento general de protección de datos

2.1.- Decisiones automatizadas y elaboración de perfiles

Este tema será tratado de forma breve, destacando aquellos aspectos esenciales para la temática de este trabajo, ya que su análisis en profundidad fue abordado en

[21] Vid. STC 46/2001 de 15 febrero.

[22] Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

[23] Sobre los estrictos requisitos que exige su tratamiento, se pronunció el Tribunal Constitucional en su STC 76/2019, de 22 de mayo, que anuló el apartado 1 del artículo 58 bis de la Ley Orgánica 5/1985, de 19 de junio, del Régimen Electoral General que regulaba la recopilación por los partidos políticos de datos personales relativos a las opiniones políticas de los ciudadanos.

[24] El CEPD recoge el estudio de KOSINSKI, M., STILWELL, D. y GRAEPEL, T.; *"Private traits and attributes are predictable from digital records of human behaviour"*, Proceedings of the National Academy of Sciences of the United States of America, volumen 110, nº 15, pp. 5802–5805. Dicho estudio, *"combinó los «me gusta» de Facebook con información limitada procedente de encuestas y halló que los investigadores predijeron con exactitud la orientación sexual de un usuario varón en el 88 % de los casos; el origen étnico de un usuario en el 95 % de los casos; y si un usuario era cristiano o musulmán en el 82 % de los casos"*. (*Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679* del GT29, Adoptadas el 3 de octubre de 2017 y revisadas por última vez y adoptadas el 6 de febrero de 2018, p. 17).

DOI: https://doi.org/10.51915/ret.309

dos estudios precedentes a los que me remito (GARRIGA DOMÍNGUEZ, 2018 y 2021). Por otra parte, en mi opinión, el derecho a la protección de datos personales se presenta como un instrumento válido, pero insuficiente, frente a las técnicas de difusión de noticias falsas, cuya viralización se produce "a través de las redes sociales como Facebook o Twitter en lo que se ha venido a bautizar como «cascada informativa" (PAUNER CHULVI, 2018: 302) y el fenómeno de la desinformación y por ello resultaba urgente la aprobación del Reglamento de Servicios Digitales para complementar, en el ámbito de las plataformas que operan en la red, en conjunto de obligaciones y garantías para los derechos fundamentales que propiciaran un entorno en línea seguro y confiable.

El artículo 22 del RGPD, establece el derecho del interesado *a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar*. Se aplica tanto a la elaboración de perfiles como a la adopción de decisiones automatizadas, estén o no basadas en perfiles. El artículo 22 del RGPD contiene una prohibición general de tomar decisiones individuales basadas únicamente en un tratamiento automatizado, incluida la elaboración de perfiles, que produzcan efectos jurídicos o efectos significativamente similares, si bien existen excepciones a esta norma general y, dichas excepciones cuando se apliquen, exigirán la adopción de medidas específicas para garantizar los derechos y libertades del interesado, así como sus intereses legítimos[25]. La aplicabilidad del artículo 22 dependerá de que se produzcan o se puedan derivar esas consecuencias relevantes para la persona; es decir, cuando produzca efectos jurídicos que afecten al interesado (por ejemplo, afecte a sus derechos, se le deniegue una prestación, produzca efectos en un contrato en el que sea parte) o le afecte significativamente de modo similar, por lo que debe ser suficientemente importante, como por ejemplo, que se le deniegue un crédito, que le afecte a su acceso a determinados servicios, en el acceso o promoción en el empleo, etc. No obstante, en determinados casos, es posible la elaboración de perfiles o la adopción de decisiones automatizadas y su licitud dependerá de que cuenten con una base de legitimación válida y cumplan los principios relativos al tratamiento. No nos encontramos ante un derecho absoluto, estableciéndose en el RGPD una serie de excepciones: cuando la decisión esté autorizada por el Derecho de la Unión o de los Estados miembros; cuando sea necesaria para la celebración o la ejecución de un contrato entre el interesado y un responsable del tratamiento; o cuando se base en el consentimiento explícito del interesado. En estos dos últimos supuestos el art. 22.3 exige que el responsable del tratamiento establezca medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado y se le garantice: el derecho a

[25] *Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679* del GT29, Adoptadas el 3 de octubre de 2017 y revisadas por última vez y adoptadas el 6 de febrero de 2018, p. 16.

DOI: https://doi.org/10.51915/ret.309

obtener intervención humana por parte del responsable; el derecho a expresar su punto de vista y el derecho a impugnar la decisión.

Finalmente, en el art. 22 se establece una limitación en razón de la naturaleza de los datos personales prohibiéndose la adopción de decisiones automatizadas basadas en datos sensibles o especialmente protegidos (origen étnico o racial, opiniones políticas, convicciones religiosas o filosóficas, afiliación sindical, datos genéticos, datos biométricos, datos relativos a la salud o datos relativos a la vida sexual o las orientación sexuales) salvo que el interesado haya prestado su consentimiento explícito y esta posibilidad no esté prohibida por el Derecho de la Unión o de los Estados miembros o cuando el tratamiento sea necesario por razones de un interés público esencial, sobre la base del Derecho de la Unión o de los Estados miembros, que debe ser proporcional al objetivo perseguido, respetar en lo esencial el derecho a la protección de datos y establecer medidas adecuadas y específicas para proteger los intereses y derechos fundamentales del interesado. En ambos casos deberán tomarse medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado.

2.2.- La importancia de los principios de transparencia y de responsabilidad proactiva del RGPD en los sistemas de IA

El RGPD exige que en cualquier tratamiento de datos personales se respeten tanto los principios relativos al tratamiento, como que este cuente con una base de legitimación adecuada[26]. El artículo 5 del RGPD regula los principios básicos que deberán respetarse en cualquier tratamiento de datos personales. Sin embargo, aunque todos ellos deben exigirse, en el contexto de los sistemas de IA que adoptan decisiones, contenidos o recomendaciones, con fines de publicidad o comerciales u otros diferentes, ocupan un papel destacado dos de ellos: el principio de transparencia y el de responsabilidad proactiva.

El principio de responsabilidad proactiva se encuentra desarrollado en el artículo 24 del RGPD, al establecer la obligación general del responsable del tratamiento de aplicar las medidas técnicas y organizativas apropiadas a fin de garantizar y poder

[26] Sobre la inadecuación de la letra b) del artículo 6.1 del RGPD (ejecución de un contrato) como base de legitimación para la publicidad de comportamiento en línea se ha pronunciado el CEPD en sus Directrices 2/2019 sobre el tratamiento de datos personales en virtud del artículo 6, apartado 1, letra b), del RGPD en el contexto de la prestación de servicios en línea a los interesados, p. 16 y ss. Señala el CEPD que, "como regla general, el tratamiento de datos personales con fines de publicidad del comportamiento no puede considerarse necesario para la ejecución de un contrato de servicios en línea" ya que resulta muy difícil demostrar "que el contrato no pueda ejecutarse debido a la ausencia de anuncios publicitarios de comportamiento". Por otra parte, "el artículo 6, apartado 1, letra b), no puede servir de fundamento jurídico para la publicidad del comportamiento en línea por el mero hecho de que este tipo de publicidad financie de manera indirecta la prestación del servicio. Aunque este tipo de tratamiento puede respaldar la prestación del servicio, ello no basta de por sí para determinar que resulta necesario para ejecutar el contrato en cuestión".

Revista Española de la Transparencia
RET. ISSN 2444-2607
Núm. 17. Número Extraordinario 2023
www.revistatransparencia.com

Págs. 137-164
DOI: https://doi.org/10.51915/ret.309

demostrar que el tratamiento es conforme con el RGPD, teniendo en cuenta la naturaleza, el ámbito, el contexto y los fines del tratamiento así como los riesgos de diversa probabilidad y gravedad para los derechos y libertades de las personas físicas. Supone un cambio de paradigma que exige pasar de un sistema de protección reactivo frente al incumplimiento a un modelo preventivo y proactivo (LORENZO CABRERA, 2018: 123-125). Va a exigir un enfoque desde el riesgo para determinar la probabilidad y la gravedad del riesgo para los derechos y libertades del interesado[27]. Será necesario realizar una evaluación objetiva atendiendo a la naturaleza, el alcance, el contexto y los fines del tratamiento de datos en el proceso de micro-segmentación. Por lo tanto, de forma previa a la focalización habrá de valorarse, por aquellos que intervengan en las diferentes fases del procedimiento, si dicho tratamiento se va a servir de datos especialmente protegidos o si va a implicar un *"alto riesgo"* para las personas y sus derechos, que exijan la realización de una EIPD (Evaluación de Impacto relativa a la Protección de Datos). Este sería el caso de anuncios dirigidos a personas vulnerables, ya que es posible que surjan "riesgos adicionales en función de los fines de la campaña publicitaria y su carácter intrusivo, o de si la focalización implica el tratamiento de datos personales observados, inferidos o derivados"[28].

Por otra parte, como una de las características de estos procesos es, que suelen ser opacos para las personas, o a éstas, les cuesta comprender su funcionamiento y la relevancia de sus aplicaciones y consecuencias, el cumplimiento del principio de transparencia resulta capital. En el RGPD, las referencias al principio de transparencia son constantes garantizando que, de forma sencilla, fácilmente accesible y en un lenguaje claro y sencillo, se facilite al interesado toda la información relevante para él en el proceso de tratamiento de sus datos. Esta obligación es especialmente pertinente *"en situaciones en las que la proliferación de agentes y la complejidad tecnológica de la práctica hagan que sea difícil para el interesado saber y comprender si se están recogiendo, por quién y con qué finalidad, datos personales que le conciernen, como es en el caso de la publicidad en línea"* (Considerando 58).

La exigencia de transparencia debe ser configurada como *"un derecho prestacional que requiere una actuación positiva por parte de las autoridades públicas" (TOMÁS MALLÉN, 2015: 832 y ss.)* y se conecta con el establecimiento de un contenido pormenorizado del derecho de información y de las correlativas obligaciones informadoras del responsable del tratamiento. Así se recoge en el artículo 13 del

[27] Este enfoque desde el riesgo es también el adoptado en la Propuesta de Reglamento de Inteligencia Artificial. La Propuesta clasifica los sistemas de IA en cuatro niveles atendiendo a su posible riesgo, en función de la gravedad del daño y de su probabilidad.

[28] Directrices 8/2020 sobre la focalización de los usuarios de medios sociales, Versión 2.0, adoptadas el 13 de abril de 2021, p. 34.

DOI: https://doi.org/10.51915/ret.309

RGPD[29], que obliga al responsable del tratamiento a adoptar las medidas oportunas para facilitar al interesado toda información relevante relativa al tratamiento de sus datos personales incluida la existencia de decisiones automatizadas y la elaboración de perfiles, así como *"información significativa sobre la lógica aplicada, la importancia y las consecuencias previstas de dicho tratamiento para el interesado"*. Igualmente, el principio de transparencia obliga al responsable del tratamiento a garantizar que se le informa, aún cuando los datos no se hayan obtenido directamente del interesado, en los términos previstos en el artículo 14 y a garantizar el derecho de acceso, en el artículo 15, a la información relativa a la existencia de decisiones automatizadas, incluida la elaboración de perfiles y, al menos en tales casos, información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento.

Por último, es necesario destacar que la obligación de transparencia se configura como "una expresión del principio de lealtad en relación con el tratamiento de los datos personales plasmado en el artículo 8 de la Carta de Derechos Fundamentales de la Unión Europea"[30]; pues, cualquier tratamiento de datos personales deberá de ser lícito y leal de forma que al interesado le ha de quedar totalmente claro que se están recogiendo y utilizando sus datos y en la medida en que éstos son o serán tratados de forma que, como señala el Comité Europeo de Protección de Datos, las personas no se vean sorprendidas "en un momento posterior del uso que se ha dado a sus datos personales"[31]. Señala COTINO, asimismo, el principio de transparencia deberá aplicarse a los datos de entrenamiento y también a los datos de entrada para su funcionamiento, validación y prueba; pues, "se trata de un elemento clave que incide directamente en la calidad y robustez del sistema, así como respecto de la posibilidad de controlar sesgos, errores o posibles discriminaciones" (COTINO HUESO, 2023: 33).

Al igual que ocurre con el enfoque desde el riesgo, las exigencias de transparencia también se recogen en la Propuesta de Reglamento sobre IA, estableciéndose "normas armonizadas de transparencia aplicables a los sistemas de IA destinados a interactuar con personas físicas, los sistemas de reconocimiento de emociones y los

[29] El incumplimiento del principio de transparencia, entre otros requisitos, ha llevado a la *Irish Data Protection Authority* a sancionar a Meta el pasado enero. Sus dos decisiones sancionadoras fueron el resultado de investigaciones basadas en reclamaciones sobre las actividades de Facebook e Instagram, en particular en relación con la legalidad y la transparencia del tratamiento para la publicidad basada en el comportamiento. La Autoridad irlandesa impuso a Meta una multa de 210 millones de euros en la decisión sobre Facebook y de 180 millones de euros en la decisión sobre Instagram. El comunicado, *"Facebook and Instagram decisions: "Important impact on use of personal data for behavioural advertising"*, puede consultarse en la página oficial del Comité Europeo de Protección de Datos (EDPD, por sus siglas en inglés) del día 12 de enero de 2023 (edpb.europa.eu/news).

[30] Directrices sobre la transparencia en virtud del Reglamento (UE) 2016/679, Adoptadas el 29 de noviembre de 2017. Revisadas por última vez y adoptadas el 11 de abril de 2018, p. 5.

[31] *Ibídem*, p. 8.

DOI: https://doi.org/10.51915/ret.309

sistemas de categorización biométrica, así como a los sistemas de IA usados para generar o manipular imágenes, archivos de audio o vídeos" (artículo 1) y exigiendo para los sistemas de IA de Alto Riesgo que se diseñen y desarrollen "de un modo que garantice que funcionan con un nivel de transparencia suficiente para que los usuarios interpreten y usen correctamente su información de salida" (artículo13). Estas exigencias de transparencia en la IA se indican también en el Libro Blanco sobre la inteligencia artificial de la Comisión, de 19 de febrero de 2020, ya que "la falta de transparencia (opacidad de la IA) hace difícil detectar y demostrar los posibles incumplimientos de la legislación, especialmente las disposiciones legales que protegen los derechos fundamentales, imputan responsabilidades y permiten reclamar una indemnización"[32].

3.- Obligaciones específicas para las plataformas en línea y motores de búsqueda de muy gran tamaño[33] en el Reglamento (UE) 2022/2065, de 19 de octubre, de Servicios Digitales (RSD). Especial referencia a las obligaciones de transparencia algorítmica

3.1.- Un Reglamento para crear un entorno en línea seguro y frenar la difusión de desinformación y manipulación en línea

El desarrollo de Internet y la proliferación de servicios en la red supuso el incremento del número y clase de servicios disponibles: acceso a toda clase de información y documentación en cualquier lugar del mundo, posibilidad de comunicarse y relacionarse con un número ilimitado de personas, acceso a nuevas formas de ocio, posibilidad de adquirir servicios o bienes de cualquier clase a través del comercio electrónico, etc. Su desarrollo modificó hábitos y costumbres de ocio, laborales y profesionales, como consumidores o en nuestras relaciones como ciudadanos con las diferentes administraciones. Internet ha transformado también las formas de actuación de las empresas y de las entidades públicas e incluso puede afirmarse que ha supuesto la desaparición de los conceptos de espacio y tiempo en las comunicaciones (CAMPUZANO TOMÉ, 2002: 17). Se desarrollaron nuevos servicios de comunicación e información, aunque, al mismo tiempo, se amplificaron los riesgos y ataques para los derechos de las personas. Internet permite fácilmente a los usuarios elaborar informaciones, opiniones y contenidos de toda clase que pueden compartir con otros grupos de personas, divulgándolos a escala planetaria. Por ello, brinda un gran potencial para la promoción de la democracia y la diversidad cultural y el ejercicio de las libertades de expresión e información. Sin embargo, existe una necesidad objetiva de controlar la red, pues es utilizada también para la

[32] Anteriormente, en la Comunicación de la Comisión de 8 de abril de 2019 "Generar confianza en la inteligencia artificial centrada en el ser humano", se identificaba este concepto como uno de los siete requisitos esenciales para lograr una IA fiable.

[33] VLOP y VLOSE por sus siglas en inglés (*Very Large Online Platforms* y *Very Large Online Search Engines).*

DOI: https://doi.org/10.51915/ret.309

comisión de delitos, como mensajero implacable de la difamación y el acoso, en especial a mujeres y grupos vulnerables o minorías, para atentar contra la vida privada de las personas, supone riesgos específicos para los derechos de la infancia, facilita los ilícitos contra la propiedad intelectual, etc.

Entre los objetivos del Reglamento de Servicios Digitales destaca el de crear un entorno en línea seguro, predecible y confiable, en el que las personas puedan ejercer sus derechos fundamentales, en particular, la libertad de expresión e información. Así se recoge en el Considerando 3 y a lo largo de toda su parte expositiva. El legislador europeo es muy consciente del papel que juega la publicidad en línea en el discurso político y en otros ámbitos sensibles como puede ser el de la salud pública o en la propagación del discurso del odio, señalando que la publicidad en línea "puede contribuir a generar riesgos significativos, desde anuncios publicitarios que sean en sí mismos contenidos ilícitos hasta contribuir a incentivar económicamente la publicación o amplificación de contenidos y actividades en línea que sean ilícitos o de otro modo nocivos"[34]. Como ya se ha señalado, la aplicación de la IA en estos ámbitos permite el perfilado ideológico individual y, a través de las técnicas de focalización, podrá elaborarse información política personalizada. De esta forma, la cantidad y calidad de la información personal que se encuentra en las redes sociales, permite a los anunciantes mejorar el alcance e impacto de su publicidad al dirigirse a grupos específicamente seleccionados y estructurados o, incluso, a individuos concretos para influir en su conducta (BARBU, BARBU, 2014: 46) y, como ha señalado el CEPD, esta práctica tendrá consecuencias negativas para el pluralismo político y el debate público de ideas, pero existirán otros riesgos, de discriminación utilizando, o no, informaciones sensibles y, asimismo, un riesgo real de manipulación, que podría afectar a cuestiones y procesos políticos, acentuando vulnerabilidades y emociones negativas[35].

Son varios los instrumentos que el RSD utiliza para lograr estos objetivos, desde la prohibición de la publicidad dirigida a menores, hasta una serie de medidas para garantizar a los usuarios un mayor control sobre el uso de sus datos personales. Refuerza el principio de transparencia y establece obligaciones para los prestadores de los servicios de informar sobre cuándo y en nombre de quién se presenta la publicidad, así como los principales parámetros utilizados para determinar que se les va a presentar publicidad específica, que ofrezca explicaciones reveladoras de la lógica utilizada con ese fin, también cuando se base en la elaboración de perfiles. Además de estas obligaciones generales de transparencia, se establecen obligaciones específicas para las plataformas en línea y buscadores de muy gran

[34] Considerando 52.

[35] Directrices 8/2020 sobre la focalización de los usuarios de medios sociales, Versión 2.0, adoptadas el 13 de abril de 2021, p. 7.

Revista Española de la Transparencia
RET. ISSN 2444-2607
Núm. 17. Número Extraordinario 2023
www.revistatransparencia.com

Sección Estudios y Artículos
Las exigencias de transparencia para los sistemas algorítmicos…
por Ana Garriga Domínguez
Págs. 137-164
DOI: https://doi.org/10.51915/ret.309

tamaño, con un número de destinatarios activos en la Unión de 45 millones o superior de promedio mensual. La razón del establecimiento de este conjunto de obligaciones particulares está clara: su relevante papel dado su alcance expresado en el número de destinatario de servicio, "para facilitar el debate público, (…) la difusión de información, opiniones e ideas y para influir en la forma en que los destinatarios obtienen y comunican información en línea"[36].

En el Capítulo III se regulan las obligaciones de diligencia debida para crear un entorno en línea transparente y seguro y se establecen las obligaciones adicionales para las plataformas en línea y las exigencias específicas para aquellas de muy gran tamaño. En los artículos 23 y siguientes se dispone la obligación de informar sobre el uso de medios automáticos con fines de moderación de contenidos. También se refuerzan, en los términos que veremos seguidamente, los requisitos de transparencia sobre la publicidad en línea, con especial atención a la transparencia de sus algoritmos.

Otra novedad, siguiendo la estela del RGPD, es el enfoque desde el riesgo que se concretará, para las plataformas de muy gran tamaño, en la obligación de evaluación de los riesgos sistémicos. Entre estos riegos sistémicos, habrán de considerarse necesariamente los riesgos de difusión de contenido ilícito a través de sus servicios o cualquier efecto negativo para el ejercicio de los derechos fundamentales a la vida privada y familiar, la libertad de expresión e información, la prohibición de la discriminación y los derechos del niño. Así mismo, deberán evaluarse los riesgos de manipulación deliberada de su servicio que pueda producir un *"un efecto negativo real o previsible sobre la protección de la salud pública, los menores, el discurso cívico o efectos reales o previsibles relacionados con procesos electorales y con la seguridad pública"* (artículo 26). Estos riesgos podrían derivarse, como se indica en el Considerando 58, de la creación de cuentas falsas, del uso de *bots* u otros comportamientos, total o parcialmente automatizados, con el resultado de una difusión rápida y extendida de información que constituya un contenido ilícito o incompatible con las condiciones de una plataforma. El RSD es consciente de que el fenómeno de la desinformación, la difusión de contenidos ilícitos o noticias falsas es complejo y en su propagación intervienen, no solo las noticias falsas, sino también las cuentas falsas y *bots,* que sirven para amplificar "la actividad e intensidad de los servicios" (LANIER, 2018: 77). Pues, como ya se ha indicado, las empresas que producen estos contenidos buscan el máximo beneficio a corto plazo para atraer el mayor número de «clics» y persiguen la viralización de la noticia y el aumento del tráfico en la red "porque eso es lo que impulsa la influencia y los ingresos por publicidad" (HOLIDAY, 2013: 290-291). Este fenómeno se ve potenciado porque su

[36] Considerando 53.

DOI: https://doi.org/10.51915/ret.309

contenido puede ser retransmitido entre los usuarios sin necesidad de un filtrado de verificación de hechos o juicio editorial significativo por parte de terceros.

Precisamente, para reducir estos riesgos se establecen refuerzan las obligaciones de diligencia[37], que impliquen la adopción de medidas correctoras adecuadas, que pueden implicar mejoras en el diseño y en el funcionamiento de los sistemas de moderación, de los sistemas algorítmicos de recomendación e de las interfaces en línea con el fin de salvaguardar el orden público y los derechos de las personas. También se prohíbe manipular las elecciones de los usuarios mediante los denominados *patrones oscuros,* en los términos que veremos seguidamente.

3.2.- Obligaciones de transparencia para los motores de búsqueda (VLOSE) y las plataformas en línea de muy gran tamaño (VLOP). Especial referencia a las obligaciones de transparencia algorítmica

El RSD propone un modelo de regulación para delimitar "el alcance de las libertades de empresa e informativas de las plataformas", que "parte de las acciones y autorregulaciones ya adoptadas por las plataformas contra la desinformación y las reorienta e impulsa hacia los objetivos necesarios" (COTINO HUESO, 2022: 236). Dentro del conjunto de obligaciones de diligencia debida para crear un entorno en línea transparente y seguro del Capítulo III del RSD, destacan las obligaciones de transparencia informativa de los prestadores de servicios intermediarios y de alojamiento de datos, excluyendo a aquellos que sean microempresas o pequeñas empresas. Estas exigencias de transparencia se aplican de forma particular a las actividades de moderación de contenidos realizada por iniciativa propia del prestador de servicios en la red incluyendo los sistemas automatizados de moderación. En concreto, el artículo 15 dispone que estos servicios habrán de publicar como mínimo una vez al año, informes claros y fácilmente comprensibles sobre cualquier actividad de moderación de contenidos que hayan realizado durante el período pertinente. Estos informes deberán incluir, en particular, la siguiente información:

- El número de órdenes recibidas de las autoridades de los Estados miembros, categorizadas según el tipo de contenido ilícito de que se trate, el Estado miembro que haya dictado la orden y el tiempo medio necesario para informar a dicha autoridad;

- Para los prestadores de servicios de alojamiento de datos, el número de notificaciones enviadas de conformidad con el artículo 16, según el tipo de contenido

[37] Obligaciones de diligencia para los prestadores de servicios de alojamiento de datos, en orden a prevenir o detectar contenidos y actividades ilegales, ya existían en la Directiva 2000/31/CE y así lo venían exigiendo los tribunales españoles en aplicación de la ley 34/2002, de 11 julio 2002, de Servicios de la Sociedad de la Información y Comercio Electrónico. Por todas, sentencia de la Sala 1ª del Tribunal Supremo 773/2009, de 9 de diciembre.

DOI: https://doi.org/10.51915/ret.309

presuntamente ilícito, el número de notificaciones enviadas por alertadores fiables, así como las actuaciones resultantes de dichas notificaciones, el número de notificaciones tratadas únicamente por medios automatizados y el tiempo medio necesario para adoptar medidas;

- Para los prestadores de servicios intermediarios, se exige que se recoja la información significativa y comprensible sobre la actividad de moderación de contenidos realizada por iniciativa propia del prestador, incluido el uso de herramientas automatizadas. Asimismo, el número y el tipo de medidas adoptadas para formar y asistir a los moderadores del servicio, así como aquellas medidas "que afecten a la disponibilidad, visibilidad y accesibilidad de la información proporcionada por los destinatarios del servicio y a la capacidad de los destinatarios para proporcionar información a través del servicio, y otras restricciones conexas del servicio";

- En el caso de los prestadores de servicios intermediarios, el número de reclamaciones recibidas a través de sus sistemas internos de gestión de reclamaciones y, para los prestadores de plataformas en línea, la base de dichas reclamaciones, las decisiones adoptadas en relación con dichas reclamaciones, con especificación del tiempo medio necesario para adoptar dichas decisiones, así como el número de decisiones revocadas;

- Cuando se utilicen medios automatizados con fines de moderación de contenidos, deberá señalarse este hecho, incluyendo una descripción cualitativa, con referencia a los fines concretos y específicos del sistema, los indicadores de la precisión y la posible tasa de error de dichos medios automatizados, así como, las salvaguardias aplicadas.

Como se ve, se trata de un conjunto de medidas que pretenden garantizar la rendición de cuentas de los prestadores del servicio y de alerta del uso de sistemas automatizados de decisión.

Estas obligaciones se complementan con las establecidas en la Sección 3 para las plataformas en línea[38]. Para el tema objeto de este trabajo, destacan dos: la prohibición del uso de patrones oscuros del artículo 25 y la especificación del contenido de la información que se debe facilitar al destinatario del servicio cuando se presente publicidad en línea, del artículo 26.

Si bien, la prohibición de los patrones oscuros, es decir, aquellas "interfaces de usuario diseñadas para influir, a través de manipulaciones psicológicas y de forma

[38] Las obligaciones de transparencia informativa del artículo 15 se completan para los prestadores de plataformas en línea en el artículo 24 que exige información adicional sobre el número, los resultados, porcentaje y tiempo medio para los litigios sometidos a los órganos de resolución extrajudicial de litigios a que se refiere el artículo 21 y el número de suspensiones impuestas en virtud del artículo 23.

DOI: https://doi.org/10.51915/ret.309

encubierta, en las elecciones del interesado, al menos, con relación al tratamiento de sus datos personales"[39], se entendía una exigencia derivada de los principios de lealtad y transparencia del artículo 5.1.a) del RGPD[40], el legislador europeo ha considerado necesario incluir la prohibición los prestadores de servicios de plataformas en línea de diseñar, organizar o gestionar "sus interfaces en línea de manera que engañen o manipulen a los destinatarios del servicio[41] o de manera que distorsionen u obstaculicen sustancialmente de otro modo la capacidad de los destinatarios de su servicio de tomar decisiones libres e informadas"[42], ya que los patrones oscuros se utilizan para que el usuario no proteja adecuadamente su privacidad, pero pueden emplearse también con otros fines.

Respecto de las obligaciones de transparencia en relación con la publicidad en línea, se exige a las plataformas que, por cada anuncio publicitario concreto presentado a cada destinatario específico, se deberá proporcionar a los destinatarios del servicio determinada información para que "sean capaces de identificar, de manera clara, concisa e inequívoca y en tiempo real", que la información es un anuncio publicitario; en nombre de quien se presenta el anuncio y quien ha pagado el anuncio, si es distinta de la aquella y la información sobre los principales parámetros utilizados para determinar el destinatario a quien se presenta el anuncio publicitario y acerca de cómo cambiar esos parámetros.

En este caso, las medidas de transparencia responden a la necesidad de clarificar el funcionamiento opaco y, generalmente, poco comprensible de cómo se realiza esta clase de publicidad, de cómo se produce la focalización y la selección del momento en que esta se ofrece. Además, en el RSD se presta especial atención a la publicidad segmentada diseñada en base a los intereses y vulnerabilidades de los usuarios por

---

[39] Guía de Protección de Datos por Defecto de la Agencia Española de Protección de Datos (AEPD), p. 20.

[40] Así lo ha entendido la AEPD (Vid. *Dark patterns: Manipulación en los servicios de Internet*. 2022). Más ampliamente, de acuerdo con la definición del CEPD, los "patrones oscuros" se consideran interfaces y experiencias de usuario implementadas en plataformas de medios sociales que llevan a los usuarios a tomar decisiones no intencionadas, involuntarias y potencialmente perjudiciales en relación con el tratamiento de sus datos personales. Los patrones oscuros pretenden influir en el comportamiento de los usuarios y pueden obstaculizar su capacidad para proteger eficazmente sus datos personales y tomar decisiones conscientes. En *Guidelines 03/2022 on deceptive design patterns in social media platform interfaces: how to recognise and avoid them.* En especial p. 3 y ss.

[41] Un conocido caso de manipulación emocional fue el que se realizó a lo largo de una semana durante el año 2012 en Facebook. En KRAMER, A., GUILLORY, J. E. y HANCOCK, J. T. 2014. *"Experimental evidence of massive-scale emotional contagion through social networks",* Proceedings of the National Academy of Sciences of Unites States of America, vol. 11, nº 24.

[42] Ejemplos reales de cómo las plataformas sociales y otros operadores utilizan el diseño de sus interfaces para manipular o engañar al usuario, alguno de ello durante sancionado por los tribunales, pueden consultarse en LACORT, J. 2018. *"Los 'dark patterns' del diseño: así hackean tu cerebro webs y aplicaciones para sacarte más dinero o retener tu atención".* Xataka, 2018. La información completa puede consultarse en la web de la revista (xataka.com) y fue publicada el día 15 de mayo de 2018.

DOI: https://doi.org/10.51915/ret.309

sus potenciales efectos negativos, prohibiéndose la elaboración de perfiles, en el sentido del Reglamento General de Protección de Datos, basada en categorías de datos especialmente protegidos de acuerdo con lo dispuesto en su artículo 9.1.

Estas obligaciones se completan en la Sección 5 con medidas adicionales de transparencia para las VLOP y los VLOSE en el artículo 39. A estos operadores se les obliga, cuando presenten publicidad en sus interfaces a recopilar y hacer públicas en una sección específica de su interfaz, a través de una herramienta de búsqueda fiable que permita realizar consultas en función de múltiples criterios, y mediante interfaces de programación de aplicaciones, un repositorio que deberá estar disponible durante "todo el tiempo en el que presenten un anuncio y hasta un año después de la última vez que se presente el anuncio en sus interfaces en línea", debiendo asegurarse "de que el repositorio no contenga ningún dato personal de los destinatarios del servicio a quienes se haya o se pueda haber presentado el anuncio y harán todos los esfuerzos que resulten razonables para garantizar que la información sea exacta y completa". En dicho repositorio se deberá recoger la siguiente información: el contenido del anuncio publicitario, incluidos el nombre del producto, servicio o marca y el objeto del anuncio; la persona física o jurídica en cuyo nombre se presenta el anuncio publicitario y la que lo haya pagado cuando sea diferente de la anterior; el período de tiempo durante el que se haya presentado el anuncio; cuando el anuncio esté destinado a presentarse en particular a uno o varios grupos concretos de destinatarios, deberá informarse de los parámetros principales utilizados para tal fin, incluidos, aquellos utilizados para excluir a uno o más de esos grupos concretos; las comunicaciones comerciales publicadas en las plataformas en línea de muy gran tamaño, que deberán estar identificadas de acuerdo con lo recogido en artículo 26.2 y, finalmente, el número total de destinatarios del servicio alcanzados, desglosado por Estado miembro para el grupo o grupos de destinatarios a quienes el anuncio estuviera específicamente dirigido.

Las obligaciones de transparencia algorítmica se extienden asimismo a los sistemas de recomendación debiendo las plataformas informar, de forma sencilla y comprensible, acerca de cuales son "los parámetros principales utilizados en sus sistemas de recomendación, así como cualquier opción a disposición de los destinatarios del servicio para modificar o influir en dichos parámetros principales". Como mínimo, el artículo 27 determina que se deberán proporcionar al destinatario del servicio, con el fin de explicarle por qué se le recomienda una determinada información, "los criterios más significativos a la hora de determinar la información sugerida al destinatario del servicio", y "las razones de la importancia relativa de dichos parámetros". Las obligaciones de transparencia para los sistemas de recomendación, se completan para las VLOP y los VLOSE en el artículo 38. Estos prestadores deberán incluir en sus sistemas de recomendación al menos una opción, para cada uno de sus sistemas, que no se base en la elaboración de perfiles en el sentido del artículo 4 del RGPD.

DOI: https://doi.org/10.51915/ret.309

Para los motores de búsqueda y las plataformas en línea de muy gran tamaño, el RSD impone obligaciones adicionales de gestión de riesgos sistémicos que atañen, tanto a la necesidad de realizar una evaluación de riesgos, "que se derive del diseño o del funcionamiento de su servicio y los sistemas relacionados con este, incluidos los sistemas algorítmicos, o del uso que se haga de sus servicios", como a la aplicación de medidas de reducción de riesgos razonables, proporcionadas y efectivas para dichos riesgos sistémicos (artículo 35).

Entre los riesgos sistémicos, el artículo 35 determina que necesariamente deberán incluir las evaluaciones se encuentran los siguientes:

- la difusión de contenido ilícito a través de sus servicios;

- cualquier efecto negativo real o previsible para el ejercicio de los derechos fundamentales, en particular los relativos a la dignidad humana, al respeto de la vida privada y familiar, a la protección de los datos de carácter personal, a la libertad de expresión e información, incluida la libertad y el pluralismo de los medios de comunicación, a la no discriminación, a los derechos del niño y a un nivel elevado de protección de los consumidores;

- cualquier efecto negativo real o previsible sobre el discurso cívico y los procesos electorales o sobre la seguridad pública;

- cualquier efecto negativo real o previsible en relación con la violencia de género, la protección de la salud pública y los menores, así como las consecuencias negativas graves para el bienestar físico y mental de la persona.

Entre los factores que imperiosamente habrán de tener en cuenta en sus evaluaciones se encuentra la necesidad de prestar atención al diseño de sus sistemas de recomendación y de cualquier otro sistema algorítmico pertinente, a sus sistemas de moderación de contenidos a las prácticas del prestador relacionadas con los datos, a los sistemas de selección y presentación de anuncios y, finalmente, a las condiciones generales aplicables y su ejecución. Asimismo, será necesario que se analice en qué medida los riesgos sistémicos pueden verse influidos y de qué manera, por "la manipulación intencionada de su servicio, en particular por medio del uso no auténtico o la explotación automatizada del servicio, así como la amplificación y la difusión potencialmente rápida y amplia de contenido ilícito y de información incompatible con sus condiciones generales" (apartado 3 del artículo 34). Si el algoritmo es el que nos indica qué es lo que queremos realmente, qué necesitamos, a quién queremos seguir o qué noticias son las que nos interesan y

DOI: https://doi.org/10.51915/ret.309

atañen, es lógico que el RSD imponga obligaciones de evaluación de riesgo desde el diseño, en especial, ante la posibilidad de manipulación intencionada[43].

Para evaluar el grado de cumplimiento de las obligaciones establecidas en el Capítulo III del RSD, así como las derivadas de los códigos de conducta en los que sean parte a las VOLP y los VLOSE habrán de someterse a auditorías anuales a su costa.

Por último, se refuerza la transparencia confiriendo, al coordinador de servicios digitales y a la Comisión, poderes de escrutinio y acceso a los datos necesarios para hacer un seguimiento de cumplimiento de las obligaciones sistémicas previstas en el mismo. Para ayudar en las funciones de ejecución y supervisión de la Comisión, se ha creado el European Centre for Algorithmic Transparency (ECAT). Entre sus funciones principales se encuentran la de análisis de la transparencia, evaluación de riesgos y propuesta de nuevos enfoques transparentes y mejores prácticas en este ámbito con el fin de mejorar la comprensión de cómo funcionan los algoritmos. Para lograr este objetivo realizará, tanto tareas de evaluación y supervisión, como de investigación y prevención. Le corresponderá trabajar para identificar los riesgos sistémicos asociados a los VLOP y VLOSE, para la propuesta de metodologías para garantizar algorítmicos transparentes y responsables o el estudio del impacto social a largo plazo de los algoritmos[44].

4.- Conclusión

El fenómeno de la desinformación no es nuevo, las teorías de la conspiración, los rumores sin verificar y los bulos ya existían antes de Internet y del desarrollo de las plataformas sociales. Sin embargo, la arquitectura y el modelo de negocio de las plataformas en línea facilita la diseminación de contenidos ilícitos de forma exponencial multiplicando su daño. El propio diseño de los sistemas de IA y de los algoritmos de los sistemas automáticos de recomendación, de selección de contenidos o de publicidad en línea explotan vulnerabilidades y permiten un extenso conocimiento sobre los intereses, gustos, opiniones ideológicas y estados emocionales de los destinatarios de los servicios. Se utilizan filtros burbuja, que

[43] El riesgo de manipulación intencionada ya fue señalado, entre otros, por el CEPD, pues, aprovechando determinados momentos en los que el análisis de la información revele estados emocionales determinados, se pueden *"dirigir a la persona mensajes específicos y en momentos concretos a los que se espera que sea más receptiva e influir así subrepticiamente en su proceso de pensamiento, sus emociones y su comportamiento"* (Directrices 8/2020 sobre la focalización de los usuarios de medios sociales 2021, 7.). Por otra parte, el reconocimiento de emociones es posible, en especial con el uso de tecnologías biométricas y pueden utilizarse para manipular a las personas. Vid. Dictamen conjunto 5/2021 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) del 18 de junio de 2021.

[44] Pueden consultarse sus funciones y misión en European Centre for Algorithmic Transparency (disponible en: algorithmic-transparency.ec.europa.eu).

Revista Española de la Transparencia
RET. ISSN 2444-2607
Núm. 17. Número Extraordinario 2023
www.revistatransparencia.com

Págs. 137-164
DOI: https://doi.org/10.51915/ret.309

explotan el sesgo de confirmación, limitando la capacidad para conocer y comprender la realidad y los puntos de vista de aquellos que, por disentir de nuestra opinión, son relegados por un sistema de inteligencia artificial. Estos mismos algoritmos para maximizar su eficiencia refuerzan las emociones negativas que son más productivas para incrementar el tiempo que pasamos conectados y captar nuestra atención[45].

Los procesos y los sistemas tecnológicos que confluyen en este fenómeno tan complejo son de difícil compresión para el usuario medio y su funcionamiento es poco transparente. La propia lógica del diseño del modelo de negocio, que se basa en la recogida masiva de datos personales para ser reelaborados en el ámbito de la realización de perfiles predictivos, que buscan condicionar el comportamiento de los individuos con diversos fines publicitarios y de marketing, son mucho más eficientes si este es un proceso es opaco. Y, si bien es cierto que los sucesivos escándalos relacionados con la influencia del fenómeno de la desinformación a través de las redes sociales han tenido como consecuencia que estos servicios adopten determinadas medidas correctoras[46], éstas se han mostrado claramente insuficientes. Por ello, resultaba urgente que se establecieran por el legislador europeo obligaciones concretas de transparencia y de rendición de cuentas para los operadores que explotan los procesos de perfilado que permiten condicionar la conducta de las personas, limitar sus opciones informativas o potenciar la desinformación. Cuando cualquier decisión se adopta con base en un perfil, incluida la información a la que tendremos acceso, el principio de transparencia es imprescindible para poder conocer quién diseña esas categorías en las que se nos clasifica, quién decide su significado y quién decide bajo qué circunstancias esas categorías serán decisivas (LYON, 2014: 186).

En la medida en que en los servicios de los operadores en línea, no es el usuario quien elige cómo relacionarse con los demás, sino que es "el proveedor del servicio el que a través del ejercicio de *default power* determina a su antojo los detalles de ese mundo compartido" (IPPOLITA, 2012: 143), el RSD impone claras obligaciones y exige un elevado grado de responsabilidad a "los actores de ecosistema (digital) que

[45] Por su negativa influencia en la salud mental de niños y adolescentes, el distrito de la escuela pública de Seattle presentó una demanda judicial colectiva contras las principales empresas de redes sociales por crear aplicaciones que explotan sus cerebros en maduración, a fin de maximizar cuánto tiempo pasan con sus plataformas y aumentar las ganancias.

(https://www.lavanguardia.com/vida/20230111/8674436/escuela-publica-seattle-demanda-redes-sociales-atacar-salud-mental-jovenes.html). Consultado el 12 de enero de 2023.

[46] Como el Código de buenas prácticas en materia de desinformación, entre cuyos firmantes estarán Google, Facebook, Twitter, Mozilla y asociaciones empresariales que representan al sector de la publicidad.

DOI: https://doi.org/10.51915/ret.309

se benefician de las conductas nocivas"[47]. Sus previsiones, incrementando las exigencias de transparencia y prohibiendo determinados comportamientos *online* se presentan como instrumentos idóneos para luchar contra el fenómeno de la desinformación. El establecimiento de obligaciones específicas para las plataformas en línea y especialmente aquellas de muy gran tamaño, así como la previsión de elevadas sanciones para las conductas más graves (artículo 52) y la garantía del derecho a presentar una reclamación de los destinatarios del servicio (artículo 53), contribuirán a proteger los derechos fundamentales de las personas en su actividad en las plataformas sociales, contribuyendo al sostenimiento de los valores democráticos. Por ello, el conjunto de medidas incluidas en el RSD para exigir a los prestadores de servicios que actúen de forma responsable y diligente para conseguir un entorno en línea seguro, predecible y que genere confianza, se deben valorar favorablemente.

No obstante, urge que el ordenamiento europeo se vea completado con la aprobación del Reglamento de IA, pues "una sola aplicación de Inteligencia Artificial puede impactar en una gran cantidad de derechos" (DE ASÍS ROIG, 2022: 102). Es necesario preservar adecuadamente los derechos de las personas que pudieran resultar afectados por los sistemas de alto riesgo y para, definitivamente, prohibir la puntuación o crédito social de las personas, los sistemas de IA que las clasifiquen a partir de sus datos biométricos en grupos por razón de su origen étnico, sexo, orientación política o sexual, así como la inferencia de emociones, que puedan dar lugar a la explotación de sus vulnerabilidades. Pues, como han señalado el SEPD y el CEPD, la IA ampliará la cantidad de predicciones que pueden hacerse, empezando por las correlaciones mensurables entre datos invisibles para los ojos humanos, pero visibles para las máquinas, lo que facilitara nuestras vidas y resolverá un gran número de problemas; pero, *"al mismo tiempo, erosionará nuestra capacidad de dar una interpretación causal a los resultados, de modo que los conceptos de transparencia, control humano, rendición de cuentas y responsabilidad por los resultados se verán seriamente cuestionados*"[48].

[47] Dictamen del SEPD sobre la manipulación en línea y los datos personales. DOUE del 4 de julio de 2018.

[48] Dictamen conjunto 5/2021 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial), p. 6-7.

DOI: https://doi.org/10.51915/ret.309

Bibliografía

ALCARÁZ RAMOS, M. 2020. "Preguntas de la explosión tecnológica del conocimiento a la política democrática y al derecho", en FUENTES SORIANO, O. (Dir.). 2020. Era Digital, Sociedad y Derecho. 2020. Valencia: Tirant lo Blanch.

ALCOTT, H. y GENTZKOW, M. 2017, *"Social Media and Fake News in the 2016 Election"*, Journal of Economic Perspectives, vol. 31, nº 2.

AÑÓN ROIG, M. J. 2022. *"Desigualdades algorítmicas. Conductas de alto riesgo para los derechos humanos"*. Derechos y libertades: Revista de Filosofía del Derecho y derechos humanos, nº 47, 2022.

BAKSHY, MESSING y ADAMIC. 2015. *"Exposure to ideologically diverse news and opinion on Facebook"*. Science, Junio de 2015, Vol. 348, nº 6239.

BARBU, BARBU, O. 2014. "Advertising, Microtargeting and Social Media", Procedia - Social and Behavioral Sciences 163.

BELLOSO MARTÍN, N. 2022. "La problemática de los sesgos algorítmicos (con especial referencia a los de género) ¿Hacia un derecho a la protección frente a los sesgos? En LLANO ALONSO, F. H. 2022. Inteligencia artificial y Filosofía del Derecho. Murcia: Laborum ediciones.

BELTRÁN PARDO, M y SEVILLANO JAÉN, F. 2013. Cloud computing, tecnología y negocio, Madrid: Paraninfo.

BENDER, E. M., GEBRU, T., MCMILLAN-MAJOR, A., & SHMITCHELL, S. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, March 2021.

BYUNG-CHUL H. 2014. La sociedad de la transparencia. Barcelona: Herder, p. 109.

CAMPUZANO TOMÉ, Herminia: *Vida privada y datos personales. (Su protección jurídica frente a la sociedad de la información),* Tecnos, Madrid, 2002.

CARDON, D. 2018. Con qué sueñan los algoritmos: nuestras vidas en el tiempo de los Big Data. Madrid: Dado Ediciones.

COTINO HUESO, L. 2022. "Quién, cómo y qué regular (o no regular) frente a la desinformación". Teoría y Realidad Constitucional. n.º 49.

DOI: https://doi.org/10.51915/ret.309

2023. "Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida". Revista Española de la Transparencia. Núm. 16. Primer semestre. Enero-junio de 2023.

CRAIG, T. y LUDLOFF, M. E. 2011. Privacy and Big Data, Sebastopol: O'Really, p. 6.

DE ASÍS ROIG, R. 2022. Derechos y tecnologías. Madrid: Dykinson.

ECHEVERRÍA, J. 2013. Entre cavernas. De Platón al cerebro pasando por Internet. Madrid: Triacastela.

GARCÍA MAHAMUT, R. 2015. *"Partidos políticos y derecho a la protección de datos en campaña electoral: tensiones y conflictos en el ordenamiento español".* UNED. Teoría y Realidad Constitucional, núm. 35, 2015.

GARRIGA DOMÍNGUEZ, A. 2015. Nuevos retos para la protección de los datos personales. En la Era del Big Data y de la computación ubicua. Madrid: Dykinson.

2018. "La elaboración de perfiles y su impacto en los derechos fundamentales. Una primera aproximación a su regulación en el Reglamento General de Protección de Datos de la Unión Europea", en Derechos y libertades: Revista de Filosofía del Derecho y derechos humanos, nº 38.

2021. "Decisiones automatizadas basadas en algoritmos y protección de datos personales", en Tomás Mallén, B., García Mahamut, R. y Pauner Chulvi, C. (Ed.): Las cláusulas específicas del Reglamento General de Protección de Datos en el ordenamiento jurídico español. Cuestiones clave de orden nacional y europeo. Valencia: Tirant lo Blanch.

HOLIDAY, R. 2013. Confía e mi, estoy mintiendo. Confesiones de un manipulador de los medios. Barcelona: Empresa Activa.

IPPOLITA. 2012. En el acuario de Facebook. El irresistible ascenso del anarco-capitalismo. Madrid: En clave de libros.

KERR, I. y EARLE, J. 2013. "Prediction, preemption, presumption: how Big Data threatens big picture privacy", Stanford Law Review, Volume 66, num. 65.

KIRLEY E. 2016. *"The robot as cub reporter: law's emerging role in cognitive journalism"*, en European Journal of Law and Technology, Vol 7, No 3.

LANIER, J. 2018. Diez razones para borrar tus redes sociales de inmediato, Barcelona: Editorial Debate.

LORENZO CABRERA, S. 2018. "Posición jurídica de los intervinientes en el tratamiento de datos personales. Medidas de cumplimiento", en AA.VV.: Protección de datos, responsabilidad activa y técnicas de garantía. Madrid: REUS.

LYON, D.2014. Surveillance Studies. An overview. Malden: Polity Press.

MAYER-SCHÖNBERGER, V. y CUKIER, K. 2013. Big Data. La revolución de los datos masivos. Madrid: Turner Noema.

O'NEIL, C. 2017. Armas de destrucción matemática. Cómo el Big Data aumenta la desigualdad y amenaza la democracia. Madrid: Capitán Swing.

PARISIER, E. 2017. El filtro burbuja: Cómo la web decide lo que leemos y lo que pensamos. Madrid: Taurus.

PAUNER CHULVI, C. 2018. "noticias falsas y libertad de expresión e información. El control de los contenidos informativos en la red". UNED. Teoría y Realidad Constitucional, núm. 41, 2018.

RAMOS BERNAL, A. 2012. Reflexiones sobre economía cuántica, Alicante: ECU.

ROLLNERT LIERN. 2002. *"La libertad ideológica en la jurisprudencia del Tribunal Constitucional".* Cuadernos y Debates, nº 129. Madrid: Centro de Estudios Políticos y Constitucionales.

SAMUEL, A. L. 1983.First grade TEX: a beginner's TEX manual. Stanford Department of Computer Science.

SARIGOL, E., GARCÍA, D. y SCHWEITZER, F. 2014. "Online Privacy as a Collective Phenomenon", Proceedings of the second edition of the ACM conference on Online social networks, ACM, octubre de 2014. http://arxiv.org/pdf/1409.6197.pdf.

TENE, O. y POLONETSKY, J. 2012. "Privacy in the age of Big Data: a time for big decisions". Stanford Law Review Online, nº 63.

TOMÁS MALLÉN, B. 2015. *"Transparencia y protección de datos: nuevos desafíos para la garantía europea de los derechos fundamentales",* en RALLO LOMBARTE, A. y GARCÍA MAHAMUT, R. 2015. Hacia un nuevo Derecho europeo de protección de datos. Valencia: Tirant lo Blanch.

XIOL RÍOS, J. A. 2001. *"La libertad ideológica o libertad de conciencia"*, en AA. VV. 2001. La libertad ideológica. Actas e las VI Jornadas de la Asociación de Letrados del Tribunal Constitucional, Cuadernos y Debates n.º 115. Madrid: Centro de Estudios Políticos y Constitucionales.

ZUBOFF, S. 2019. La era del capitalismo de la vigilancia. Barcelona: Paidós.