

La discriminación algorítmica en España: límites y potencial del marco legal

Un programa de



GOBIERNO
DE ESPAÑA

VICEPRESIDENCIA
PRIMERA DE GOBIERNO

MINISTERIO
DE ASUNTOS ECONÓMICOS
Y TRANSFORMACIÓN DIGITAL

SECRETARÍA DE ESTADO
DE DIGITALIZACIÓN
E INTELIGENCIA ARTIFICIAL

red.es



MOBILE
WORLD CAPITAL
BARCELONA

Sobre

Digital Future Society

Digital Future Society es una iniciativa transnacional sin ánimo de lucro que conecta a responsables políticos, organizaciones cívicas, expertos académicos y empresarios para explorar, experimentar y explicar cómo se pueden diseñar, usar y gobernar las tecnologías a fin de crear las condiciones adecuadas para una sociedad más inclusiva y equitativa.

Nuestro objetivo es ayudar a los responsables políticos a identificar, comprender y priorizar los desafíos y las oportunidades fundamentales, ahora y en los próximos diez años, en relación con temas clave que incluyen la innovación pública, la confianza digital y el crecimiento equitativo.

Para más información, visite digitalfuturesociety.com

Un programa de



Permiso para compartir

Esta publicación está protegida por la licencia internacional Creative Commons Attribution-ShareAlike 4.0 (CC BY-SA 4.0).

Publicado

Septiembre del 2022

Aviso legal

La información y las opiniones expuestas en este informe no reflejan necesariamente la opinión oficial de Mobile World Capital Foundation. La Fundación no garantiza la exactitud de los datos incluidos en este informe. Ni la Fundación ni ninguna persona que actúe en nombre de la Fundación será considerada responsable del uso que pueda darse a la información que contiene.

Contenidos

| | |
|--|-----------|
| 1. Introducción | 4 |
| 2. La discriminación algorítmica | 6 |
| La discriminación algorítmica y los desafíos éticos del uso de la inteligencia artificial | 6 |
| Características de la discriminación algorítmica | 9 |
| La protección de datos, una herramienta insuficiente frente a la discriminación algorítmica | 12 |
| 3. Retos que la discriminación algorítmica plantea al derecho antidiscriminatorio europeo y español | 14 |
| Retos relativos a las categorías del derecho antidiscriminatorio: discriminación directa e indirecta | 15 |
| Problemas relativos a la interseccionalidad y la granularidad de los datos | 17 |
| Obstáculos en el abordaje práctico de los casos de discriminación algorítmica | 18 |
| Dificultades en cuanto a la atribución de responsabilidad en la discriminación algorítmica | 19 |
| 4. Opciones de intervención | 20 |
| Aplicación de mecanismos de transparencia | 20 |
| Regulación de los procesos de decisión automatizados o semiautomatizados | 22 |
| Desarrollo de políticas contra la discriminación algorítmica | 24 |
| Aplicación de soluciones tecnológicas a los sesgos algorítmicos | 26 |
| 5. Conclusiones y recomendaciones | 28 |
| Referencias | 31 |
| Agradecimientos | 35 |

1. Introducción

A medida que el uso de aplicaciones basadas en inteligencia artificial (IA) se va extendiendo en todos los ámbitos de nuestra vida personal y de las sociedades en las que vivimos, van creciendo también las llamadas de atención sobre estas tecnologías.

El uso de la IA plantea numerosos desafíos éticos, además de los tecnológicos. Uno que está suscitando cada vez más preocupación es el de la discriminación algorítmica o los resultados discriminatorios que puede producir el uso de sistemas de decisión automatizada o semiautomatizada, especialmente, pero también de otras aplicaciones basadas en la IA.

Con un marco regulatorio todavía en fase de desarrollo, es importante conseguir identificar qué es la discriminación algorítmica y cómo podemos hacerle frente. De momento, la mayoría de los casos que saltan a la prensa y al debate público están relacionados con resultados discriminatorios producidos por el uso de datos personales sensibles, como el sexo o la pertenencia étnica o racial. Sin embargo, las normativas de protección de datos personales no abarcan todos los criterios ni los mecanismos de la discriminación, y tampoco establecen medidas que puedan solucionar estos casos.

En el desarrollo de las normas que regulen y controlen el uso de la IA, la discriminación necesita un tratamiento específico, que se conecte con los mecanismos de garantía antidiscriminatorios ya existentes y previstos por la Constitución española y el derecho. Ello evitaría áreas de exención respecto al principio de igualdad y la proliferación de sistemas basados en la IA a los que se permite ignorar o violar derechos fundamentales y el principio de igualdad.

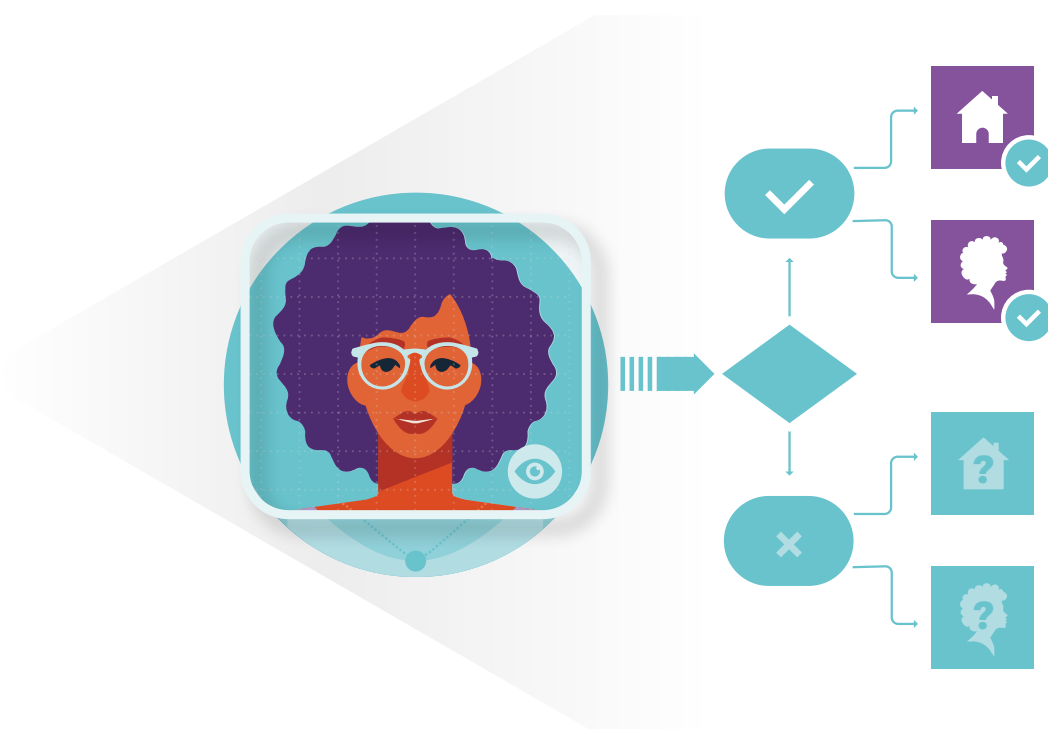
Este informe se incluye en un debate más amplio y forma parte del trabajo que realiza Digital Future Society Think Tank para explorar los retos a los que se enfrenta el sector público al implementar sistemas automatizados de toma de decisiones. Es continuación, por tanto, del trabajo que hemos efectuado para conocer cómo esos sistemas pueden discriminar a ciertos colectivos y perpetuar y exacerbar sesgos de nuestras sociedades (Digital Future Society 2020 a,b; Digital Future Society 2021).

En particular, el informe nace de la necesidad de entender las implicaciones legales de estos sistemas, y de saber si las leyes antidiscriminatorias y de igualdad de género en España pueden manejar adecuadamente los daños algorítmicos que puedan surgir. Para hacerlo, se sitúa la discriminación algorítmica en el centro del análisis. También las divergencias entre tecnólogos, juristas y analistas de políticas públicas a la hora de entender el problema, y las dificultades que pueden resultar de ello.

Se analizan, además, los desafíos que la discriminación algorítmica plantean al derecho antidiscriminatorio y a su doctrina, y se muestran algunas de las debilidades del marco actual para ofrecer protección frente a las características de la discriminación algorítmica.

El informe propone recomendaciones tanto para el desarrollo del marco regulatorio de la IA como para el de una política pública antidiscriminatoria contra la discriminación algorítmica. La discriminación algorítmica no genera un problema nuevo, que no existiese, pero tampoco es meramente un error de unos sistemas todavía no suficientemente desarrollados y que puede tener una solución tan solo tecnológica.

La discriminación algorítmica resulta de la interacción compleja entre el uso de las tecnologías de la información y la comunicación (TIC) y la IA y las dinámicas sociales ya existentes de desigualdad, los estereotipos y las jerarquías sociales. Abordar la discriminación algorítmica significa abordar esta complejidad.



2. La discriminación algorítmica

La discriminación algorítmica y los desafíos éticos del uso de la inteligencia artificial

Las decisiones automatizadas o semiautomatizadas son cada vez más frecuentes. Tanto las Administraciones públicas como las empresas privadas usan la IA y los algoritmos para apoyar o generar decisiones en diversas áreas. Las soluciones basadas en la IA se aplican a asuntos tan diversos como la música que escuchamos, los anuncios que nos aparecen en el ordenador o el móvil o la información que recibimos a través de ellos, las oportunidades de empleo a las que accedemos, la evaluación de nuestra solvencia crediticia, la credibilidad que se otorga a nuestras denuncias o el cálculo del riesgo de que seamos víctimas de un delito.

La IA promete decisiones neutrales, eficientes, basadas en datos, más rápidas y de menor coste respecto a las tomadas tras los tradicionales procesos realizados por seres humanos. Ante la expansión de la IA y, especialmente, ante el papel de la IA y los algoritmos en la toma de decisiones, se pueden observar actitudes opuestas. Las hay entusiasmadas, que han sido tildadas de *fetichismo tecnológico*, y, por el contrario, también una desconfianza y un escepticismo crecientes debidos a la opacidad de muchos de estos sistemas y las dificultades técnicas para entender cómo funcionan.

Un número creciente de estudios muestran que los sistemas de decisión automatizada o semiautomatizada basados en algoritmos, tanto en el ámbito público como en el privado, pueden afectar a los derechos fundamentales.

Veamos uno de los casos estudiados: en febrero del 2020, un tribunal de La Haya (Países Bajos)¹ declaró ilegal un sistema algorítmico utilizado por el Gobierno neerlandés, denominado SyRI (Systeem Risico Indicatie). El objetivo del algoritmo consistía en prevenir y combatir el fraude contra la seguridad social y la hacienda pública, y su funcionamiento se basaba en el uso del *big data* en la elaboración de perfiles que, asignados a cada ciudadano o ciudadana, determinaban el nivel de riesgo y la probabilidad de fraude para con las arcas públicas.

El tribunal sentenció que SyRI no cumplía con las exigencias de proporcionalidad y transparencia necesarias, y que infringía el derecho a la vida privada establecido en el artículo 8.2 del Convenio Europeo de Derechos Humanos de las personas, mayoritariamente, de renta baja y origen inmigrante, que vivían en “distritos problemáticos” (Eguiluz Castañeira 2020).

Otro caso: en diciembre del 2020, un tribunal de Bolonia (Italia) sentenció que el denominado algoritmo Frank, usado por Deliveroo para categorizar (*rankear*) a sus *riders* (repartidores), discriminaba. El algoritmo ordenaba la categoría “índice de confiabilidad” en base a las faltas y las no cancelaciones a tiempo de sus turnos dentro del sistema de reserva de vacantes para realizar los encargos (*slots*), sin valorar las causas de fondo de las ausencias de los repartidores. En resumidas cuentas, para el sistema de IA era lo mismo faltar al trabajo por desobediencia o por razones banales que por una enfermedad o por el ejercicio del derecho a huelga (Oliva León 2021).

Según el tribunal, el hecho de que el algoritmo no tenga en cuenta las razones que hay detrás de una cancelación equivale a una discriminación y penaliza injustamente a los trabajadores con razones legalmente legítimas para no trabajar (Ibid).

La preocupación política por los desafíos sociales que plantea el uso de la IA se puede ver en la publicación de normas y guías que intentan canalizar el fenómeno: desde el Libro Blanco de la Comisión Europea² hasta la reciente Propuesta sobre la IA (Ley de IA)³. También en los anuncios sobre la creación de organismos de vigilancia como el Comité Europeo de IA⁴ o la Agencia Española de Supervisión de la IA (AESIA)⁵.

Los desafíos a la protección de los derechos fundamentales que plantea el uso de las nuevas TIC y de sistemas de IA se han estudiado especialmente en relación con la protección de los datos personales y la privacidad, pero abarcan un abanico mucho más amplio (Agencia de los Derechos Fundamentales de la Unión Europea y Consejo de Europa 2020).

¹ Sentencia del Tribunal de Distrito de La Haya, de 5 de febrero de 2020: <https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878>

² Comisión Europea. Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza, del 19 de febrero de 2020: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf

³ Comisión Europea. 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) y se modifican determinados actos legislativos de la Unión, del 21 de abril de 2021: <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:52021PC0206>

⁴ Artículos 56-58 de la Propuesta de Reglamento sobre la IA.

⁵ Pérez, Enrique. Los algoritmos de las redes sociales serán controlados por una agencia pública: el Gobierno vigilará cómo se aplica la IA, del 19 de noviembre de 2021: <https://www.xataka.com/legislacion-y-derechos/algoritmos-redes-sociales-seran-controlados-agencia-publica-gobierno-quiere-vigilar-como-se-aplica-ia>

La utilización incontrolada de la IA puede poner en jaque la democracia y los valores sobre los que se fundamenta el Estado de derecho. Es lo que sucedió en el caso de las llamadas *cámaras de eco* producidas por los algoritmos de las redes sociales, que afecta a la libertad de información y al derecho a voto (Battaglini 2019), o en el de la amplificación de los bulos y mensajes de odio que atentan contra la dignidad humana y el derecho al honor.

Asimismo, las nuevas generaciones de herramientas algorítmicas ofrecen oportunidades sin precedentes para la vigilancia indiscriminada o masiva, el llamado *capitalismo de vigilancia* (Zuboff 2020), la elaboración de perfiles y la clasificación de los ciudadanos.

Además, los sistemas de IA pueden afectar a la tutela judicial efectiva y a las garantías procesales. Y es que, debido a su opacidad o falta de transparencia, no permite conocer los procesos internos con los que los algoritmos generan sus resultados. De esa manera, suscitan incertidumbre a la hora de determinar la aplicabilidad y ejecución de las disposiciones legales que protegen los derechos fundamentales, atribuyen responsabilidades y permiten reclamar indemnizaciones.

Junto a la acción política, se han incrementado también las contribuciones científicas en el ámbito que estamos analizando, a través de la constitución de foros interdisciplinarios (TIC, ingeniería, derecho, ética) e intersectoriales (empresas tecnológicas, grupos de investigación y organizaciones de la sociedad civil).

La mayor parte de la atención académica se ha centrado en la violación de derechos (como el derecho a la privacidad o algunos principios generales del derecho penal) que puede derivar del uso, el mal uso o el uso fraudulento de las tecnologías. Aun así, el problema específico de la discriminación algorítmica empieza a reclamar atención y abrirse paso.

Hasta ahora, la investigación sobre discriminación y algoritmos se ha centrado, sobre todo, en el contexto estadounidense (Barocas y Selbst 2016; Gillis y Spiess 2019), pero empiezan a publicarse también análisis del fenómeno en Europa (Hacker 2018; Zuiderveen Borgesius 2020; Xenidis y Senden 2020).

La relación entre la IA y la discriminación, y el impacto que la IA puede tener sobre la desigualdad, el principio de igualdad y el de no discriminación, han sido objeto de estudios encargados tanto por el Consejo de Europa (Zuiderveen Borgesius 2018) como por la Comisión Europea (Gerards y Xenidis 2020), además de otros realizados por agencias internacionales y nacionales (Orwat 2020; Défenseur des droits y CNIL 2020; Balayn y Gürses 2021).

Características de la discriminación algorítmica

Según el Grupo independiente de Expertos de Alto Nivel sobre la IA, la discriminación algorítmica, sesgo de la IA o sesgo algorítmico consiste en la existencia de errores sistemáticos y repetibles en un sistema informático que conllevan o generan resultados (*outputs*) discriminatorios o injustos desde un punto de vista legal, como favorecer a un grupo de usuarios en detrimento de otros (Grupo independiente de Expertos de Alto Nivel sobre la Inteligencia Artificial 2020).

Hay varios tipos de sesgos (NIST 2021) que desempeñan un papel relevante en el contexto de la IA:

- **El sesgo de automatización:** una confianza excesiva en los resultados o soluciones generados por los algoritmos atenúa las habilidades humanas.
- **El sesgo heredado:** la utilización de datos de entrenamiento sesgados como entradas en un sistema de IA reproduce el sesgo, lo que afectará al resultado futuro (por ejemplo, un algoritmo de cribado de currículums que se nutre de datos con un sesgo de género y aprende que las mujeres son menos válidas para un puesto de trabajo).
- **El sesgo de correlación:** la correlación de diferentes conjuntos de datos por un algoritmo puede desembocar en prácticas discriminatorias (por ejemplo, asociar el género a una menor productividad en el trabajo o rendimiento laboral, no por una relación causal, sino porque históricamente las mujeres han sido evaluadas más negativamente que los hombres por el mismo rendimiento laboral).
- **El sesgo de amplificación:** la capacidad de los algoritmos para procesar cantidades masivas de datos a gran velocidad crea un riesgo específico consistente en la “elevación de escalas”, que puede afectar a números muy elevados de personas en tiempos demasiado breves como para que noten los problemas o los puedan mitigar. Esto implica, por tanto, la exacerbación y aceleración de la reproducción de los estereotipos y sesgos discriminatorios que reflejan los datos con los que se alimenta el sistema de IA.

La discriminación y los sesgos algorítmicos se pueden asociar a diversos factores.

El primero de estos factores es la **calidad de los datos**. La utilización de datos incompletos, sesgados, incorrectos u obsoletos se considera una fuente primaria de sesgos algorítmicos. A veces se llega a argumentar que los algoritmos, en realidad, no discriminan: no pueden hacerlo porque no son más que matemáticas, solo recopilan y procesan datos. Si el resultado es discriminatorio, hay que mirar a los datos: un argumento que se ha expresado gráficamente como “basura entra, basura sale” (Xenidis y Senden 2020, p. 157).

El impacto diferenciado en base a motivos de discriminación prohibidos como el género, la raza u origen étnico, la discapacidad y la edad que muestran diversos estudios (Barocas y Selbst 2016; Eubanks 2018) se atribuye entonces a las desigualdades estructurales incrustadas en los datos utilizados por los algoritmos.

La calidad de los datos puede producir sesgos en varios momentos: las muestras de entrenamiento utilizadas para los sistemas de aprendizaje automático (*machine learning*) pueden estar sesgadas, o los datos a los que el sistema tiene acceso pueden reflejar jerarquías sociales arraigadas, representaciones incorrectas o insuficientes de ciertos grupos sociales, una distribución desigual de bienes, oportunidades y cargas, etc.

También podría suceder que los diseñadores y desarrolladores de modelos de IA introduzcan sus propios sesgos y prejuicios. Pueden incorporarlos voluntaria o involuntariamente cuando preparan las muestras de entrenamiento, o cuando definen el peso de cada variable. Un ejemplo ya clásico son los algoritmos de los sistemas de reconocimiento facial o de los sistemas de clasificación de género que utilizaban *datasets* de entrenamiento sesgados con una representación insuficiente de mujeres y personas racializadas y producían, por ello, resultados con una tasa de error mayor (más del 34,4%) en el caso de mujeres no blancas respecto al de hombres blancos (Hao 2019).

Además, en comparación con las decisiones sesgadas que toman los seres humanos basadas en esos mismos prejuicios y jerarquías que componen los datos de los que se nutren los algoritmos, las decisiones automatizadas o semiautomatizadas basadas en criterios discriminatorios plantean un **problema de escala**. Es decir, aplican las decisiones de manera más rápida y a un mayor número de individuos, y reducen la probabilidad de identificar y abordar los problemas a tiempo. Los algoritmos de aprendizaje automático también pueden discriminar a gran escala a través de soluciones que reproducen patrones de desigualdad y estereotipos incrustados en los datos de los que aprenden o se alimentan.

Otro factor asociado a la discriminación algorítmica se refiere a la **imprevisibilidad de los resultados**. Si el sistema de IA es demasiado extenso y complejo, puede no ser entendido por seres humanos, ni siquiera por personas expertas, y producir falta de transparencia y responsabilidad en el proceso de toma de decisiones. Es lo que ocurrió, por ejemplo, en el caso de las redes neuronales profundas modernas para datos de texto a gran escala (Buonso y Park 2020).

Los algoritmos pueden tener, además, un comportamiento parcialmente autónomo, es decir, pueden funcionar interpretando determinados datos de entrada y utilizando un conjunto de instrucciones predeterminadas, sin limitarse a ellas, a pesar de que el comportamiento del sistema esté limitado y orientado a cumplir el objetivo que se le haya asignado y otras decisiones pertinentes de diseño.⁶ En este caso, la posibilidad de entender, controlar o predecir el funcionamiento concreto del algoritmo se atenúa, incluso para los propios programadores del sistema. Con ello se complica, además, la determinación de la responsabilidad por el impacto o el resultado de la aplicación del sistema.

⁶ Parlamento Europeo. Resolución sobre inteligencia artificial: cuestiones de interpretación y de aplicación del Derecho internacional en la medida en que la UE se ve afectada en los ámbitos de los usos civil y militar, así como de la autoridad del Estado fuera del ámbito de la justicia penal: https://www.europarl.europa.eu/doceo/document/TA-9-2021-0009_ES.html

Finalmente, otro factor relativo a la discriminación algorítmica –porque dificulta su detección y tratamiento– es la **opacidad o efecto “caja negra” (black box)**, que hace difícil determinar dónde reside la raíz de la discriminación. Los sistemas de decisión automatizada o semiautomatizada pueden contener prejuicios y sesgos involuntarios de sus programadores, o, si son intencionales, estar ocultos o enmascarados en una codificación muy compleja.

Los algoritmos pueden producir resultados discriminatorios para algunos grupos, no porque usen una categoría prohibida entre sus códigos, sino a través de valores de aproximación (*proxies*) y correlaciones establecidas dentro del contexto del *big data* y, por lo tanto, muy difíciles, si no imposibles, de rastrear o revisar. La opacidad o efecto “caja negra” envuelve tanto el funcionamiento del algoritmo como, muy a menudo, la calidad de los datos que maneja, lo que dificulta ulteriormente la determinación del problema de discriminación.

El problema de la opacidad de los algoritmos se debe tanto a problemas técnicos como a cuestiones legales: la complejidad de comprender su codificación y procesamiento desde un razonamiento humano dificulta o impide la transparencia y la accesibilidad. Pero, además, los modelos y los algoritmos son productos protegidos por las leyes de propiedad intelectual y de secretos empresariales, y los desarrolladores y sus clientes no están, generalmente, dispuestos a ponerlos en abierto a disposición del público.

Este problema lo muestra claramente el caso que enfrenta a la plataforma cívica CIVIO con la Administración española (Ministerio para la Transición Ecológica). Ante la denegación por parte del Ministerio del bono social a personas que cumplían los requisitos para recibirlo, CIVIO solicitó información sobre el sistema BOSCO, que utilizan las compañías eléctricas para introducir los datos de las solicitudes y comunicar la respuesta. Los tribunales denegaron a CIVIO el acceso al código del programa, alegando que está amparado por la protección intelectual, pese a que ha sido desarrollado por la propia Administración, e impidió así que se pudieran contrastar los errores del programa y tramitar las reclamaciones.^{7,8}

La caracterización de la discriminación algorítmica como “sesgos” o errores del sistema que conllevan resultados inaceptables o injustos plantea, además, algunos problemas en relación con la definición jurídica, y con el sentido social y político, de la discriminación. Los examinaremos con más detalle en la sección sobre los retos que la discriminación algorítmica plantea al derecho antidiscriminatorio.

⁷ Para más información, véase <https://civio.es/tu-derecho-a-saber/2019/05/16/la-aplicacion-del-bono-social-del-gobierno-niega-la-ayuda-a-personas-que-tienen-derecho-a-ella/>

⁸ El caso inicial involucró al Consejo de Transparencia y Buen Gobierno (CTBG), pero, después de que CIVIO apeló la decisión, el CTBG retiró su oposición a divulgar el código fuente. El caso está en curso.

La protección de datos, una herramienta insuficiente frente a la discriminación algorítmica

La mayoría de las intervenciones referidas a los desafíos éticos en el contexto de la IA han tenido como objetivo la protección de datos personales y la privacidad. Algunos de estos casos muestran impactos discriminatorios, pero hasta la fecha han sido tratados con los instrumentos jurídicos que garantizan la protección de los datos personales y no mediante las normas antidiscriminatorias. Así ocurrió en el caso de SyRi, pues, aunque el tribunal mencionó el impacto discriminatorio del sistema, fue tratado como un caso de privacidad.

La normativa de protección de datos es una herramienta a la que, a veces, puede recurrirse para hacer frente a casos de discriminación en la aplicación de sistemas de IA, ya que las categorías de datos personales protegidos y los motivos de discriminación prohibidos coinciden en algunos casos.

Por ejemplo, la Agencia Neerlandesa de Protección de Datos obligó al Consejo de Distrito de Charlois, en enero del 2011, a dejar de procesar datos relativos a la raza para decidir medidas de orden público y a destruir ese tipo de información de sus bases de datos. Esta misma Agencia impuso, en enero del 2020, una multa de 750.000 euros a TikTok porque su información sobre privacidad estaba únicamente en inglés y no en neerlandés. Se consideró que esta práctica violaba la privacidad de las niñas y niños, considerados por la legislación como un grupo vulnerable.⁹

En ambos casos, se producía un perjuicio discriminatorio en grupos protegidos por las cláusulas antidiscriminatorias por motivos de pertenencia racial o étnica o por motivos de edad, respectivamente. Sin embargo, las decisiones no analizaban el impacto discriminatorio en sí, y la protección que reciben los grupos implicados frente a la discriminación es “solo” un beneficio indirecto de la protección de sus datos personales.

La normativa sobre protección de datos muestra varias limitaciones y plantea diversos problemas que impiden considerar la protección de datos como una estrategia suficiente en los casos de discriminación.

En primer lugar, las decisiones basadas en algoritmos **pueden producir efectos discriminatorios sin hacer uso de datos personales**. Los algoritmos **establecen patrones probabilísticos** a través de inferencias y valores de aproximación en procesamientos masivos de datos de todo tipo.

Por ejemplo, la evaluación de la solvencia crediticia de un cliente de una entidad bancaria puede resultar de la correlación establecida entre su código postal y una serie de valoraciones respecto a hábitos de consumo, impagos o patrones sobre alquileres y propiedad que el algoritmo asocia a ese código postal. Dada la segregación espacial étnica, racial y de clase social existente en las ciudades europeas, el código postal puede actuar como un valor de aproximación en relación con criterios de discriminación prohibidos (como la

⁹ Para más información, véase <https://www.globalcompliancenews.com/data-privacy/data-protection-enforcement-in-the-netherlands/>

raza o la pertenencia étnica) o criterios de discriminación social y estructural (como la clase social o el poder adquisitivo, que no se consideran prohibidos en todos los sistemas antidiscriminatorios).

En segundo lugar, la normativa sobre protección de datos **no sería tampoco aplicable** a procesos de decisión con resultados discriminatorios **cuando el tratamiento de datos personales se realiza en terceros países no sujetos al RGPD** y los datos se importan a Europa anonimizados o seudonimizados (higienizados).

Finalmente, la misma normativa de protección de datos **ha mostrado problemas de aplicación práctica debidos a su amplitud y abstracción** (Zuiderveen Borgesius 2020). El Parlamento Europeo ha apuntado a la falta de recursos y facultades de las autoridades de Protección de Datos como causa de su exigua aplicación, que ha hecho recaer sobre la ciudadanía la carga de la ejecución mediante la interposición de demandas ante los órganos jurisdiccionales.¹⁰ Es improbable, por tanto, que esta normativa tenga más que una incidencia esporádica sobre la discriminación algorítmica.

¹⁰ Parlamento Europeo. Resolución sobre el informe de evaluación de la Comisión sobre la ejecución del Reglamento General de Protección de Datos dos años después de su aplicación: https://www.europarl.europa.eu/doceo/document/B-9-2021-0211_ES.html



3. Retos que la discriminación algorítmica plantea al derecho antidiscriminatorio europeo y español

La discriminación algorítmica, o los efectos discriminatorios que pueden resultar del uso de procesos de decisión basados en algoritmos, plantea diversos retos al derecho antidiscriminatorio en Europa, y, en particular, en España.

Conviene destacar que la definición jurídica de discriminación no se ajusta a la definición técnica anteriormente aportada de discriminación algorítmica o sesgo algorítmico. La definición jurídica de discriminación no cubre tampoco todo el espectro de situaciones de desigualdad que social y políticamente se consideran discriminatorias.

La definición de discriminación algorítmica que se usa en contextos técnicos se basa en la idea de sesgo o error (en el diseño del modelo o como producto de la mala calidad de los datos), mientras que las ideas jurídicas y sociales de la discriminación se centran en la idea de desventaja injusta. Es conveniente tener en cuenta, por tanto, que el problema que se plantean tecnólogos y juristas bajo el nombre de discriminación algorítmica puede no ser el mismo problema, y por tanto, las soluciones que busquen pueden ser mutuamente irrelevantes o divergentes.

Retos relativos a las categorías del derecho antidiscriminatorio: discriminación directa e indirecta

El derecho antidiscriminatorio europeo se basa en la doble prohibición de la discriminación directa y de la discriminación indirecta. El derecho antidiscriminatorio español recoge ambas categorías en el artículo 6 de la Ley Orgánica de Igualdad,¹¹ en lo que se refiere al sexo, y en el artículo 28(1)(b) de la Ley 63/2003,¹² que traspone en el ordenamiento español las Directivas antidiscriminatorias del año 2000,¹³ en lo referido a origen racial o étnico, religión o convicciones, discapacidad, edad y orientación sexual.

Se entiende por **discriminación directa** la situación en que se encuentra una persona que sea, haya sido o pudiera ser tratada¹⁴ de manera menos favorable que otra en una situación comparable, por razón de alguna de las características prohibidas.

Como se puede ver, lo que define la discriminación directa es que el trato diferenciado está basado en uno de los motivos prohibidos. En el derecho de la Unión Europea, los motivos prohibidos son el sexo, la raza o la etnia, la religión o las creencias, la orientación sexual, la edad y la discapacidad. En el derecho español, la prohibición de discriminación en el artículo 14 de la Constitución no se refiere a una lista exhaustiva de motivos, sino que contiene lo que se conoce como una cláusula abierta.¹⁵ Los motivos de discriminación contemplados pueden ser, por ello, más que los de las Directivas europeas.

En general, se considera improbable que la categoría de discriminación directa sea aplicable o de utilidad en el caso de la discriminación algorítmica (Hacker 2018; Xenidis y Senden 2020; Gerards y Xenidis 2020). Pero no es, como se argumenta algunas veces, porque los algoritmos son solo matemáticas y no tienen prejuicios ni intención de discriminar; en Europa, a diferencia de lo que ocurre en el derecho antidiscriminatorio estadounidense, la intencionalidad no juega ningún papel a la hora de determinar la discriminación, ni siquiera la discriminación directa.

Se trata más bien de que los procesos de decisión automatizada basados en algoritmos manejan grandes cantidades de datos para generar sus resultados y es improbable que la decisión esté basada en uno de los criterios prohibidos, o que se pueda determinar que se ha

¹¹ Ley Orgánica 3/2007, de 22 de marzo, para la igualdad efectiva de mujeres y hombres: [BOE-A-2007-6115](#)

¹² Ley 63/2003, de 30 de diciembre, de medidas fiscales, administrativas y del orden social: [BOE-A-2003-23936#A27](#)

¹³ Directiva 2000/43/CE del Consejo, de 29 de junio de 2000, relativa a la aplicación del principio de igualdad de trato de las personas independientemente de su origen racial o étnico: <https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=LEGISSUM:l33114> y Directiva 2000/78/CE del Consejo, de 27 de noviembre de 2000, relativa al establecimiento de un marco general para la igualdad de trato en el empleo y la ocupación: <https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32000L0078&from=ES>

¹⁴ La Ley 63/2003 hizo una transposición defectuosa de las Directivas europeas y, a diferencia de la Ley Orgánica de Igualdad, no contempla los casos de discriminación pasada (haya sido) o hipotética (pudiera ser) en la prohibición de la discriminación directa por razón del origen racial o étnico, religión o convicciones, discapacidad, edad y orientación sexual.

¹⁵ Artículo 14 de la Constitución española: "Los españoles son iguales ante la ley, sin que pueda prevalecer discriminación alguna por razón de nacimiento, raza, sexo, religión, opinión o cualquier otra condición o circunstancia personal o social".

usado (o en qué medida) un criterio prohibido para decidir, vista la complejidad del código algorítmico y de los procesos que originan dichas decisiones (el problema de la opacidad o efecto “caja negra” del que ya hemos hablado).

Los efectos discriminatorios en el uso de algoritmos se deben frecuentemente a los sesgos en los datos que se usan tanto en la fase de “entrenamiento” del algoritmo como en las fases posteriores. Los datos con los que se entrena al algoritmo pueden no ser suficientemente representativos de ciertos grupos, como sucedió en los sistemas de reconocimiento facial, que fallaban con mujeres, personas racializadas y especialmente mujeres no blancas. La razón era que las imágenes con las que se había entrenado al sistema eran predominantemente de hombres blancos (Buolamwini y Gebru 2018).

También puede suceder que los datos a los que accede el algoritmo reflejen los estereotipos y la discriminación estructural que sufren ciertos grupos en la sociedad (por ejemplo, los datos que reflejan la segregación de las mujeres en diferentes sectores laborales o los de la brecha salarial), y el algoritmo no haga sino trabajar, reproducir y exacerbar esos sesgos.

Por todo ello, la discriminación algorítmica parece más abordable desde la perspectiva de la **discriminación indirecta**, que se fija en los efectos y no en las acciones.

La discriminación indirecta es la situación en que una disposición, criterio o práctica aparentemente neutros pone en desventaja a una persona respecto de otras por razón de su sexo, origen racial o étnico, religión o convicciones, discapacidad, edad u orientación sexual. Con la salvedad, eso sí, de que dicha disposición, criterio o práctica puedan justificarse objetivamente en atención a una finalidad legítima y que los medios para alcanzar dicha finalidad sean necesarios y adecuados.¹⁶

Para determinar un caso de discriminación indirecta no necesitamos saber, por tanto, si el algoritmo ha usado o no criterios prohibidos en el proceso de formación de la solución o de la decisión que ofrece el sistema. Podemos evitar así las dificultades derivadas tanto del efecto “caja negra” como del uso de valores de aproximación y correlaciones, o de los sesgos existentes en los datos a los que accede el algoritmo. Basta comprobar si esa solución/decisión tiene efectos diferenciados sobre alguno de los grupos protegidos por las categorías recogidas en la legislación europea o en la doctrina constitucional.

¹⁶ Además de las prohibiciones de discriminación indirecta recogidas en la Ley 63/2003 y en la Ley Orgánica 3/2007, el Tribunal Constitucional estableció en 1991 que la prohibición de discriminación en el artículo 14 de la Constitución se refiere tanto a la discriminación directa como a la indirecta. Sentencia del TC de 1 de julio de 1991: [ECLI: ES:TC:1991:145](#).

Problemas relativos a la interseccionalidad y la granularidad de los datos

Un problema que la discriminación algorítmica comparte con otros ámbitos de la discriminación, pero que aquí encontramos de manera exacerbada, es el relativo a la interseccionalidad. La interseccionalidad se refiere a las situaciones de desventaja creadas por la interacción de múltiples factores de discriminación o de desigualdad: por ejemplo, la manera diferenciada en la que experimentan la discriminación las mujeres negras respecto a las mujeres blancas o los hombres negros (Crenshaw 1989).

En el contexto de la toma de decisiones algorítmicas, y especialmente en el caso del aprendizaje automático, es poco probable que se tome una decisión sobre la base de una sola característica protegida. Por contra, las soluciones suelen tener en cuenta una multitud de factores y variables correlacionados estadísticamente en grandes volúmenes de datos (Hacker 2018, p. 1151).

Se ha observado que el riesgo es exactamente el contrario, que la discriminación se vuelva más “fina”, más “granular”, sumamente interseccional, y supere el número limitado de categorías protegidas (Xenidis y Senden 2020, p. 163). Por ejemplo, en el trabajo de Virginia Eubanks (Eubanks 2018) se muestra cómo la valoración hecha por los algoritmos da posibilidades de acceso a recursos sociales diferenciadas para las personas sin hogar, que presentan unas trayectorias específicas de relación con los servicios sociales y se localizan en áreas urbanas determinadas.

Aunque tuvieran un efecto diferenciado en hombres y mujeres, o por grupos de edad, o por pertenencia racial o étnica, estas categorías protegidas estarían disgregadas por la interacción de los demás “ejes. Eso haría muy difícil el enfoque del caso a través de la comparación con personas en las que no concurre una de las características protegidas.

Ni el derecho antidiscriminatorio europeo ni el español contienen normas específicas sobre la discriminación interseccional, ni hay guías sobre cómo tratar los casos interseccionales. Eso sí, la discusión sobre cómo abordar las dificultades que plantea la interseccionalidad en el derecho antidiscriminatorio ha ido creciendo en las últimas dos décadas. Especialmente, desde la adopción de las Directivas antidiscriminatorias del 2000, que multiplicaron los motivos de discriminación prohibidos en el derecho europeo.

Sin embargo, no se ha encontrado una solución y en la literatura especializada hay muchas dudas de que este concepto pueda resultar operativo en el marco del derecho antidiscriminatorio actual (Schiek 2016). Nos encontramos, entonces, que el marco jurídico antidiscriminatorio existente se encuentra mal preparado para afrontar una de las características salientes de los casos de discriminación en el contexto de la IA, que es, precisamente, el alto grado de interseccionalidad o la granularidad de las categorías.

Obstáculos en el abordaje práctico de los casos de discriminación algorítmica

Aunque desde un punto de vista conceptual la discriminación indirecta parece adecuarse al fenómeno de la discriminación algorítmica, la aplicación práctica de la protección antidiscriminatoria plantea diversos problemas.

En primer lugar, **puede ser difícil establecer la desventaja de un grupo protegido**. Aunque la discriminación indirecta no requiere la comparación con un grupo “en situación similar”¹⁷, es necesario identificar un grupo. Como ya hemos visto, el alto nivel de interseccionalidad o granularidad en las categorías usadas por los algoritmos para producir sus soluciones puede hacer difícil la identificación del grupo desaventajado y la representatividad de ese grupo en términos de las categorías protegidas.

Es decir, puede ser difícil percibir a un conjunto de personas como un “grupo” si el algoritmo ofrece, por ejemplo, soluciones menos ventajosas a un subgrupo muy particular de mujeres en determinados barrios o zonas de código postal, o en determinados sectores laborales o con algunas características muy específicas en términos de consumo de bienes o uso de servicios; también puede ser difícil que ese específico subgrupo se considere representativo de la categoría “sexo”.

Algunas otras dificultades prácticas **derivan de la falta de transparencia**. Es difícil que las víctimas de discriminación indirecta sean conscientes de serlo, y mucho más si su situación se debe a la aplicación de un algoritmo en un universo potencial de víctimas que no tienen continuidad o relación entre ellas. Es lo que ocurre si, por ejemplo, el algoritmo discrimina en la oferta de puestos de trabajo a través de una plataforma a mujeres o miembros de minorías étnicas o raciales, que no tienen modo de contrastar los resultados del sistema entre ellas.

En general, las víctimas de la discriminación indirecta, y especialmente cuando se trata de fenómenos estructurales de discriminación como pueden ser las soluciones basadas en sesgos de los datos, no están en la posición de entender los procesos que les ponen en una situación de desventaja, ni tienen la posibilidad de acceder a los datos ni recursos (conocimiento, tiempo y dinero) para plantear casos judiciales de discriminación.

Otro punto problemático de la discriminación indirecta en relación con la discriminación algorítmica es que, aunque se muestre la desventaja de la solución sobre un grupo protegido, **no se considerará que hay discriminación si se presenta una justificación objetiva**. Buena parte de la literatura científica considera que las oportunidades que ofrece la justificación objetiva minan la capacidad de la discriminación indirecta para tratar casos de discriminación algorítmica.

La justificación objetiva requiere que quien use un sistema de decisión basado en algoritmos muestre que persigue un fin legítimo y que el uso es apropiado y proporcional. Algunos autores (Hacker 2018) sostienen que la función predictiva de los algoritmos funcionaría, por sí misma, como un fin legítimo (por ejemplo, como mecanismo de medición de la productividad laboral o de la solidez crediticia) y aparecería como apropiada respecto a ese fin. Sería difícil para los jueces establecer que otro algoritmo o las soluciones ofrecidas por decisores humanos tienen menos sesgos o utilizan correlaciones menos discriminatorias.

¹⁷ Sentencias del TJUE de 10 de marzo de 2005, Nikoloudi: [ECLI:EU:C:2005:141](#), y de 17 de julio de 2014, Leone: [ECLI:EU:C:2014:2090](#).

Dificultades en cuanto a la atribución de responsabilidad en la discriminación algorítmica

En derecho antidiscriminatorio, tanto la discriminación directa como la discriminación indirecta deben ser imputables a un sujeto (los empresarios, la Administración, los proveedores de servicios, etc.). Tanto el derecho antidiscriminatorio europeo como el español prohíben las instrucciones para discriminar y las consideran como discriminación.

Sin embargo, en el caso de la discriminación algorítmica, la atribución de la responsabilidad, la imputabilidad de la acción discriminatoria, sea directa o indirecta, es más complicada, puesto que pueden intervenir una pluralidad de sujetos. Esto incluye desde el sistema de aprendizaje automático mismo, que actúa de forma relativamente autónoma en la producción de soluciones, hasta los desarrolladores del sistema, la empresa que ofrece servicios o productos (software) para implementar sistemas de decisión automatizadas o semiautomatizadas, y, obviamente, el cliente que compra esos servicios o productos y que finalmente implementa el sistema.

El Tribunal de Justicia de la Unión Europea (TJUE) estableció que tiene que haber un nexo entre los actos de los empleadores y los efectos discriminatorios para que se pueda atribuir la responsabilidad en un caso relativo a las diferentes condiciones salariales entre empleados y trabajadores contratados a través de una empresa de trabajo temporal.¹⁸

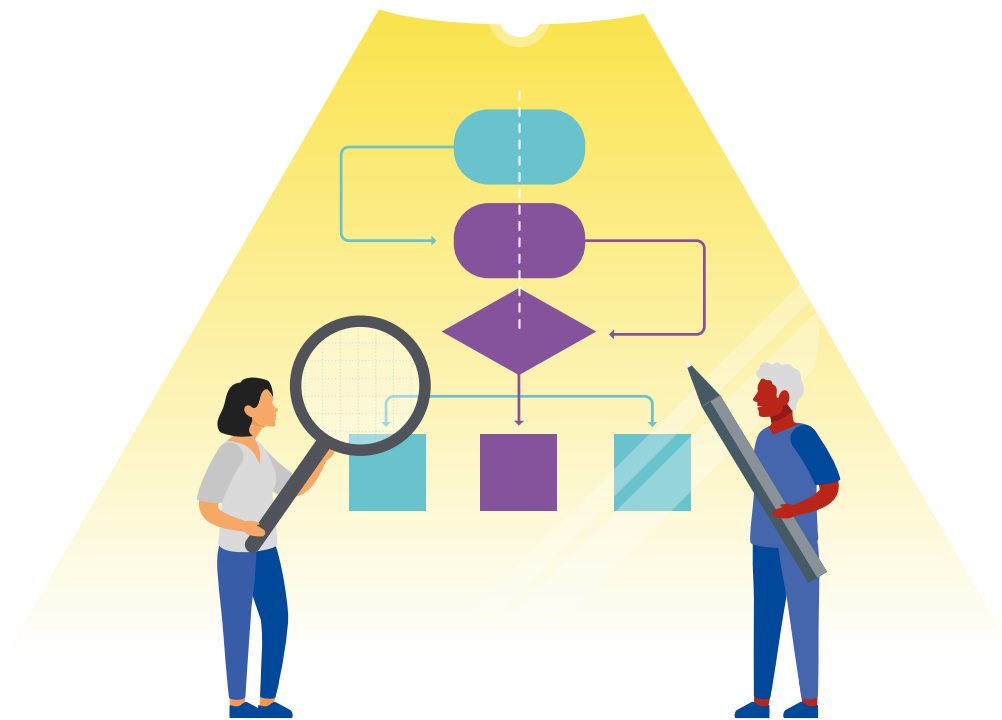
En cambio, en España, tanto la jurisprudencia como la legislación han establecido mecanismos alternativos que permiten atribuir la responsabilidad de forma presunta o incluso objetiva. El Tribunal Supremo,¹⁹ por ejemplo, ha establecido una presunción de discriminación indirecta cuando, frente a efectos discriminatorios, no haya un mínimo de transparencia en los criterios sobre los que se basó la decisión. Es lo que les podría ocurrir a los clientes que han comprado servicios o productos para implementar soluciones automatizadas y no pueden explicar en base a qué criterios se adoptaron las soluciones.

Además de la presunción de responsabilidad, los legisladores pueden optar por establecer formas de responsabilidad objetivas, es decir, que atribuyen por defecto la responsabilidad a un determinado sujeto, independientemente de la imputabilidad de la acción.

Una forma de responsabilidad objetiva la encontramos, por ejemplo, en las directrices de la Agencia Española de Protección de Datos (AEPD) del 2020. Estas establecen que cualquier persona que adopte una decisión sobre el uso de IA no puede evitar la responsabilidad con el argumento de no tener suficiente información o conocimiento técnico; la responsabilidad no puede trasladarse a los desarrolladores o a la empresa que ofrece los productos de IA.

¹⁸ Sentencia del TJUE de 13 de enero de 2004, Allonby: [ECLI:EU:C:2004:18](#).

¹⁹ Sentencia del TS de 18 de julio de 2011: [ECLI:ES:TS:2011: 5798](#).



4. Opciones de intervención

Aplicación de mecanismos de transparencia

Uno de los problemas fundamentales a la hora de identificar y combatir la discriminación algorítmica, como hemos visto, es la opacidad de los sistemas de IA y la lógica que, como propietarios, siguen las empresas que los desarrollan. Es incluso difícil saber quién y cuándo utiliza sistemas de decisión basados en algoritmos.

Por ello, una opción de intervención clara es incrementar la transparencia en relación con el uso de estas tecnologías. España tiene algunas normas destinadas a incrementar la transparencia o que podrían servir a este fin.

El Real Decreto-ley 9/2021 (la llamada *Ley Rider*) introdujo un nuevo párrafo en el artículo 64.4 del Estatuto de los Trabajadores que da al comité de empresa el derecho a “ser informado por la empresa de los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de IA que afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, acceso y mantenimiento del empleo, incluida la elaboración de perfiles”²⁰.

De esta forma, los representantes de las personas trabajadoras tendrán derecho a conocer el uso de algoritmos por parte de la empresa tanto en el caso de sistemas de decisión automatizada como cuando el algoritmo afecte a la toma de decisiones (aunque esta sea solo humana) (Monclús y Martínez 2021).

²⁰ Ley 12/2021, de 28 de septiembre, por la que se modifica el texto refundido de la Ley del Estatuto de los Trabajadores, aprobado por el Real Decreto Legislativo 2/2015, de 23 de octubre, para garantizar los derechos laborales de las personas dedicadas al reparto en el ámbito de plataformas digitales: [BOE-A-2021-15767](https://www.boe.es/BOE-A-2021-15767)

También la reciente modificación del Texto Refundido de la Ley General para la Defensa de los Consumidores y Usuarios (TRLGDCU),²¹ en su artículo 82, exige nuevos requisitos en materia de información a los consumidores y usuarios. Entre ellos destaca la información general sobre los principales parámetros que determinan la clasificación de los bienes y servicios en las búsquedas online realizadas por consumidores y usuarios. Esto abre la posibilidad de exigir jurídicamente a los empresarios información significativa sobre cómo interactúan los conjuntos de datos y los parámetros utilizados por el algoritmo.

En tercer lugar, la Ley de IA que prepara la Comisión Europea, en su artículo 13,²² exige que el diseño de los sistemas de IA de alto riesgo sea suficientemente transparente para que los usuarios puedan interpretar y utilizar correctamente sus resultados. Los sistemas de IA deberán ir acompañados de instrucciones de uso con información concisa, completa y clara. Esto incluye las características, capacidades y limitaciones del funcionamiento del sistema de IA, es decir, las especificaciones relativas a los datos de entrada, de entrenamiento, de validación, etc., así como las medidas de vigilancia humana establecidas.

Otra norma relevante, pero limitada al ámbito de los datos personales, es el artículo 13.2 f) del RGPD, que exige informar a los interesados sobre la existencia de decisiones automatizadas, incluida la elaboración de perfiles. En el caso de la elaboración de perfiles, la Ley Orgánica 3/2018 de Protección de Datos Personales y garantía de derechos digitales,²³ en su artículo 11.2, da a los afectados el derecho a ser informados y a oponerse a la adopción de decisiones individuales automatizadas que produzcan efectos jurídicos sobre ellos o que les afecten significativamente en modo similar. El alcance de esta norma está marcado por el uso de los derechos de propiedad intelectual y el secreto empresarial como límites al derecho a la información.

Finalmente, una norma relevante que podría ser fácilmente aplicada al uso de sistemas de decisión basados en algoritmos es la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno.²⁴ El artículo 12 de esta Ley da derecho de acceso a la información pública, es decir, a los contenidos o los documentos elaborados o adquiridos en el ejercicio de sus funciones por parte de las Administraciones públicas y otros sujetos sometidos al deber de transparencia, como entidades gestoras, agencias estatales o entidades públicas o participadas.

Si para la distribución de ayudas públicas o en la determinación del acceso a servicios se utilizan decisiones automatizadas, la información relativa al código y la lógica aplicada por el código deberían encontrarse sometidas a los mismos requisitos que se establecen en los artículos 6, 7 y 8 de dicha Ley para la planificación y producción normativa, así como los actos con relevancia jurídica, los que tengan repercusión económica y presupuestaria o las estadísticas necesarias para valorar el cumplimiento y calidad de los servicios públicos.

²¹ Real Decreto Legislativo 1/2007, de 16 de noviembre, por el que se aprueba el texto refundido de la Ley General para la Defensa de los Consumidores y Usuarios y otras leyes complementarias: [BOE-A-2007-20555](https://www.boe.es/boe/BOE-A-2007-20555)

²² Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial: <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex:52021PC0206>

²³ Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales: [BOE-A-2018-16673](https://www.boe.es/boe/BOE-A-2018-16673)

²⁴ Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno: [BOE-1-2013-12887](https://www.boe.es/boe/BOE-1-2013-12887)

Sin embargo, al valorar la importancia de establecer requisitos de transparencia relativos al uso de algoritmos en la toma de decisiones y a las características de estos, hay que tener en cuenta que no se trata únicamente de proveer información, sino de que se materialice en un derecho a la explicabilidad (Selbst y Powles 2017), es decir, a que esta información sea comprensible por parte de los destinatarios.

Este es un derecho controvertido (Wachter et al. 2017), y se argumenta a menudo que los códigos de muchos algoritmos no son comprensibles o trazables ni siquiera para expertos en IA. Sería aconsejable que las normas que imponen requisitos de transparencia establecieran también los mecanismos que garanticen la explicabilidad o que mitiguen su falta, cuando no es técnicamente posible trazar o entender el algoritmo.

Regulación de los procesos de decisión automatizados o semiautomatizados

La regulación explícita de los procesos de decisión automatizados o semiautomatizados con el fin de evitar la discriminación es un reto jurídico de gran calado para la Unión Europea.

La actual propuesta de Ley de IA, por ejemplo, categoriza los sistemas de IA en función de su riesgo para los derechos fundamentales y las libertades de los ciudadanos. El principio de no discriminación ocupa un lugar destacado entre las motivaciones del documento, que se aventura a considerar que: "...la propuesta complementa el Derecho de la Unión vigente en materia de no discriminación al establecer requisitos específicos que tienen por objeto reducir al mínimo el riesgo de discriminación algorítmica".²⁵

Sin embargo, cabe notar que la discriminación, como tal, no se menciona ni una sola vez en el articulado de la propuesta.

La propuesta establece un sistema de clasificación de los sistemas de IA basado en diferentes niveles de riesgo. Un sistema de IA se considera de alto riesgo en base a la función que lleva a cabo y a la finalidad específica o las modalidades para las que se usa el sistema. La propuesta contiene un Anexo III que enumera los sistemas de IA cuyos riesgos ya se han materializado y autoriza a la Comisión a ampliar esa lista.

Para los diferentes niveles de riesgo, se establecen obligaciones específicas de prevención y mitigación. Los sistemas de alto riesgo tienen obligaciones relativas al diseño de sistemas de evaluación de riesgo, a la calidad de los datos, a la documentación técnica, al registro y trazabilidad del sistema, a las garantías de supervisión humana y a la ciberseguridad y solidez del sistema.

Numerosas organizaciones han criticado este enfoque: 61 de ellas escribieron una carta abierta a la Comisión pidiendo que se establecieran "líneas rojas" en el uso de sistemas de IA y que se prohibieran aquellos usos que reproducen la discriminación estructural o suponen un atentado a los derechos fundamentales (EDRi 2021).

²⁵ Propuesta de Reglamento de la IA, p. 4.

El enfoque basado en el riesgo ha sido asimismo criticado por su falta de rigor (Mahler 2021). También porque no se dirige fundamentalmente a gestionar el riesgo, sino que busca un equilibrio, una proporcionalidad entre los riesgos que plantean los sistemas de AI y los derechos fundamentales o los valores democráticos (Access Now 2021).

El enfoque de niveles de riesgo no está pensado para hacer frente al riesgo de discriminación específicamente. Y eso a pesar de que la lista de sistemas de alto riesgo del Anexo III contiene ámbitos y funciones de decisión en los que la discriminación es habitual (el acceso a la educación o al empleo, la gestión de la migración, el acceso a bienes y servicios, etc.).

Es decir, las obligaciones que se recogen en la propuesta, para los sistemas de IA que se aplican en estos ámbitos y con funciones de clasificación, selección e incluso “predicción” (léase cálculo de probabilidades), no tienen en cuenta las características específicas de la discriminación algorítmica, como el uso de valores de aproximación, la granularidad de las categorías, los problemas relativos a la identificación de la desventaja o el problema de la justificación objetiva. Esto último es especialmente relevante vista la importancia que la propuesta atribuye a la idea de proporcionalidad.

También la Propuesta de Reglamento sobre servicios digitales (Ley de servicios digitales)²⁶ reconoce que determinados grupos o personas pueden encontrarse en situación de vulnerabilidad o desventaja en el uso de los servicios en línea por razón de su género, raza u origen étnico, religión o convicciones, discapacidad, edad u orientación sexual. Pueden verse afectados “por prejuicios (conscientes o no) que pueden incorporarse a los sistemas de notificación por los usuarios y por terceros, y reproducirse en las herramientas automatizadas de moderación de contenidos”.

Dicha propuesta pretende mitigar este riesgo de discriminación, incluyendo la presentación discriminatoria de publicidad que afecte a la igualdad de trato y oportunidades. La propuesta recoge, además, la regulación de los sistemas internos de reclamación y de resolución extrajudicial de conflictos en línea para los destinatarios del servicio. Asimismo, se establece la obligación de realizar auditorías y un mecanismo supranacional de supervisión, con especial atención a las plataformas en línea de gran tamaño.

Como se puede ver, las propuestas de normativas europeas, aunque subrayan el riesgo de discriminación y de afectación de los derechos fundamentales, no establecen en su articulado mecanismos de derecho antidiscriminatorio, y ni siquiera contienen una prohibición de discriminación indirecta o una prohibición de instrucciones discriminatorias. Lo que establecen son mecanismos que aumentan la transparencia, la trazabilidad y el acceso a la información de los sistemas de IA, así como algunas obligaciones relativas a la evaluación de riesgos y al establecimiento de planes de mitigación.

En la legislación española, hay también algunas normas que podrían contribuir a regular el riesgo de discriminación en el uso de la IA.

²⁶ Comisión Europea. Propuesta de Reglamento del Parlamento Europeo y del Consejo relativo a un mercado único de servicios digitales (Ley de servicios digitales) y por el que se modifica la Directiva 2000/31/CE: <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020PC0825&from=en>

La Ley 3/1991 de competencia desleal²⁷, por ejemplo, establece que “el tratamiento discriminatorio del consumidor en materia de precios y demás condiciones de venta se reputará desleal, a no ser que medie causa justificada”. Es decir, si un sistema automatizado de determinación de precios en un sitio web, por ejemplo, ofreciese precios más altos en función del código postal de zonas habitadas por minorías étnicas y la empresa no pudiera dar una justificación objetiva de la diferencia de precios, nos encontraríamos con un caso de discriminación prohibida y con una práctica de competencia desleal (Battaglini 2020).

La Ley General para la defensa de los consumidores y usuarios²⁸ tipifica como infracción toda actuación discriminatoria contra personas consumidoras vulnerables independientemente del motivo o contra cualquier consumidor o usuario por el ejercicio de los derechos que confiere esta Ley, las prácticas comerciales desleales, y las conductas discriminatorias en el acceso a bienes y servicios, en especial en lo relativo a la igualdad entre mujeres y hombres.

Aunque dicha ley no tipifica específicamente las relaciones mediadas por sistemas de IA, la prohibición de discriminación es lo suficientemente extensa para comprenderlas. Sin embargo, en este terreno de la discriminación en el acceso a bienes y servicios, no hay apenas jurisprudencia que pueda servir de guía. La escasez de casos judiciales sobre discriminación en este contexto –en el que los estudios muestran la incidencia del uso de sistemas automatizados o basados en la IA (Pandey and Caliskan 2021)– se debe en buena parte a las dificultades que se han comentado anteriormente de conciencia entre las víctimas de estar siendo discriminadas y de acceso a sistemas de información y apoyo para presentar reclamaciones.

Desarrollo de políticas contra la discriminación algorítmica

Un tercer ámbito de intervención para reforzar la acción antidiscriminatoria sería el desarrollo de políticas públicas de control, seguimiento y contraste del fenómeno de la discriminación algorítmica en manos de una agencia especializada (como la propuesta AESIA) o mediante mecanismos de coordinación entre agencias especializadas (como la Oficina Antidiscriminación de la Inspección de Trabajo, el Ministerio de Igualdad, etc.).

En este ámbito de intervención hay amplio margen de mejora, puesto que el derecho antidiscriminatorio en España no está, generalmente, acompañado por políticas antidiscriminatorias ni de seguimiento y control. Tampoco hay entes públicos o colaboraciones entre entes públicos y privados que ofrezcan apoyo y tutela a las víctimas de discriminación, especialmente a las víctimas de formas estructurales o sistémicas de discriminación.

²⁷ Ley 3/1991, de 10 de enero, de Competencia Desleal: [BOE-A-1991-628](#)

²⁸ Real Decreto Legislativo 1/2007, de 16 de noviembre, por el que se aprueba el texto refundido de la Ley General para la Defensa de los Consumidores y Usuarios y otras leyes complementarias: [BOE-A-2007-20555](#)

En cuanto a la primera opción, las labores de seguimiento y control, falta una política de generación de conocimiento y evidencia sobre la efectividad de la implementación de normas antidiscriminatorias en general y sobre la incidencia del fenómeno de la discriminación en sus diversas formas (tanto del impacto en grupos diferenciados de personas, como en los diferentes sectores o ámbitos en los que ocurre la discriminación).

Una opción de intervención que parece de mayor actualidad son las auditorías algorítmicas,²⁹ que son precisamente una de las medidas a implantar por la futura AESIA.

Existen ya proyectos como el de OASI, desarrollado por la empresa pionera Eticas Foundation, consistente en un buscador que permite conocer los algoritmos utilizados por Gobiernos y empresas para tomar decisiones cruciales sobre las vidas de la ciudadanía de manera automatizada. OASI informa sobre quién ha desarrollado los algoritmos, quién los usa, su ámbito de aplicación (tanto geográfico como sectorial), si han sido auditados, sus objetivos, su impacto social y las amenazas que representan (Jiménez 2021).

Asimismo, OASI se plantea la posibilidad de auditar algoritmos de empresas o instituciones sin acceder a su código, a través de la auditoría inversa. Esto es, analizando su impacto una vez se haya implementado el algoritmo, lo que permitiría un análisis acorde con la categoría de discriminación indirecta.

Además del efecto general sobre la generación de conocimiento y la posibilidad de control, las auditorías algorítmicas también pueden ser beneficiosas desde el punto de vista de las empresas o las Administraciones públicas que usan sistemas de IA:

- En relación con la **vulneración de derechos fundamentales**: conculcar derechos sistemáticamente puede acarrear responsabilidades civiles, administrativas e incluso penales, por no hablar del daño moral provocado a los ciudadanos que ven cómo se desvanecen sus esperanzas y se restringen sus derechos y libertades.
- En términos de **economía y eficiencia**: las prácticas discriminatorias pueden resultar fatales para las arcas de la Administración o de las empresas que implementen sistemas de IA, vista la escala y el impacto numérico de los procesos de decisión automatizada y, por tanto, el potencial número de víctimas que podrían reclamar indemnizaciones o compensaciones.
- En términos de **coste reputacional**: la posibilidad de verse inmiscuido como compañía o institución en un caso de discriminación algorítmica puede suponer un importante varapalo a nivel de marketing, y reducir confianza por parte de los ciudadanos y consumidores en la utilización de tecnologías.

²⁹ La auditabilidad se refiere a la capacidad de un sistema de IA para someterse a la evaluación de los algoritmos, los datos y los procesos de diseño del sistema, según la Lista de Evaluación para una IA Fiable (Grupo independiente de Expertos de Alto Nivel sobre la Inteligencia Artificial 2020).

Aplicación de soluciones tecnológicas a los sesgos algorítmicos

La mayoría de las propuestas de intervención sobre los problemas de discriminación algorítmica que aparecen en la documentación de la UE son soluciones tecnológicas. La concepción de la discriminación en términos de sesgos resultantes de la mala calidad de los datos ha propiciado este enfoque tecnológico. En este ámbito, se entiende la discriminación no tanto como el reflejo de un problema estructural de desigualdad social, sino como un problema técnico en el diseño de los modelos, en los datos de entrenamiento, en la recogida y análisis de datos, etc. Problemas que pueden, y deben, ser solventados a través de soluciones técnicas “antis sesgo” (*debiasing*).

Por mecanismos antis sesgo se entienden los métodos que confrontan los sesgos introduciendo una forma de paridad estadística (llamada métrica justa o *fairness metric*) en el conjunto de datos, en el algoritmo o en los resultados. Con este mismo enfoque trabajan también las auditorías, que se refieren a procesos de evaluación sobre la paridad estadística del sistema examinado.

Algunos trabajos apuntan ya a las limitaciones de este enfoque tecnocéntrico de la discriminación en el contexto de la IA, así como de las soluciones antis sesgo (Balayn y Gürses 2021). De las soluciones tecnológicas se critica que las indicaciones sobre qué son los sesgos y sus problemas son muy vagas, tanto en el Libro Blanco como en la propuesta de Ley de IA, y no podrán servir como guía. Se considera que los mecanismos antis sesgo no son una panacea universal para cualquier tipo de aplicación de IA y tienen, a día de hoy, una eficacia comprobada muy limitada, que la documentación europea parece ignorar (*Ibid.*).

Además, la aproximación tecnocéntrica a la discriminación como sesgo de los datos no puede dar cuenta de la discriminación estructural (que puede tener perfecta correspondencia estadística y ser, sin embargo, injusta). Los mecanismos antis sesgo no están diseñados para confrontar el problema de la discriminación en sentido social.

Finalmente, se ha notado también que las soluciones tecnológicas antis sesgo concentran aún más poder en compañías tecnológicas que son, a la postre, actores comerciales cuyo interés principal es su propio beneficio y no las importantes cuestiones políticas y sociales que subyacen en las decisiones sobre igualdad y discriminación.

Puede que la reducción de la discriminación a sesgos pueda causar una evaluación incorrecta de los problemas de discriminación, y que los mecanismos antis sesgo no sean la panacea contra la discriminación algorítmica. Pero destacar que la IA también puede contribuir a la supervisión y seguimiento del uso de sistemas de IA, o al establecimiento de evidencias sobre la discriminación y sobre la desigualdad más general, y revelar nuestros propios sesgos (Kleinberg y Sunstein 2019).

Por ejemplo, el sistema Claudette usa el aprendizaje automático para identificar cláusulas abusivas en los términos de servicio o en las informaciones sobre privacidad de las plataformas online (Lagioia y Sartor 2020). Este sistema, desarrollado a través de una colaboración entre el mundo académico y organizaciones de consumidores, muestra, además, las ventajas de expandir el uso de la tecnología hacia actores diversos de las grandes corporaciones tecnológicas (Big Tech) o de los Gobiernos. Permite así mitigar uno de los riesgos ya mencionados en el enfoque tecnocéntrico.

Por ello, como auguran la Agencia de los Derechos Fundamentales (Agencia de los Derechos Fundamentales de la Unión Europea y Consejo de Europa 2018) o el Parlamento Europeo,³⁰ el *big data* y la IA también podrían representar oportunidades y ser instrumentos en la lucha contra la discriminación y en la protección de los derechos fundamentales.

³⁰ Parlamento Europeo. Resolución con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_ES.html, considerando 27 y ss.

5. Conclusiones y recomendaciones

Entre los desafíos éticos que plantea el uso de la IA, la discriminación algorítmica empieza a atraer la atención por su capacidad para afectar a valores y derechos fundamentales de nuestras democracias, a la cohesión social y a los esfuerzos de las políticas contra la desigualdad.

La discriminación algorítmica presenta perfiles distintivos que necesitan soluciones específicas. El uso de las normativas sobre protección de datos personales parece, a todas luces, insuficiente para hacer frente a los problemas que plantea la discriminación algorítmica.

Un aspecto fundamental, de cara al diseño de soluciones antidiscriminatorias en el contexto de la IA, es la divergencia entre las diferentes concepciones de discriminación algorítmica que tienen tecnólogos, juristas y gestores de política pública. La necesaria colaboración interdisciplinar para generar soluciones a un problema tan complejo como la discriminación en el contexto de la IA requiere que esa divergencia de concepciones sea reconocida como tal y tratada. Solo así se podrá evitar que esas diferentes concepciones se utilicen en manera intercambiable, y generen confusión y soluciones no acertadas a problemas mal planteados.

Esto se vuelve especialmente urgente cuando se empiezan a producir normas para regular el uso y aplicación de la IA, como la propuesta de Ley de IA europea, que plantean el problema de la discriminación en sentido jurídico-político en su exposición de motivos y limitan después sus soluciones a los problemas de la discriminación algorítmica en sentido tecnológico.

Es necesario lograr un mayor nivel de integración interdisciplinar en el diseño de soluciones, tanto tecnológicas como legales y de política pública.

En España, el marco antidiscriminatorio presenta evidentes debilidades para hacer frente a un fenómeno como la discriminación algorítmica. Esta puede ser una oportunidad para reforzar dicho marco normativo frente a fenómenos estructurales de discriminación y desigualdad, mediante el desarrollo de una auténtica política pública antidiscriminatoria.

Es necesario que la normativa que regule la IA recoja explícitamente la prohibición de la discriminación directa, indirecta e interseccional, y que el uso de sistemas de IA, sistemas de decisión automatizada o semiautomatizada u otras aplicaciones no se convierta en una eximente frente a los resultados discriminatorios de tal uso.

Algunas características del derecho antidiscriminatorio hacen que resulte excesivamente limitado para hacerse cargo de la discriminación algorítmica. Es preciso revisar tanto los mecanismos de determinación de las desventajas como las fórmulas para resolver los problemas de operatividad de la interseccionalidad.

Esto podría suponer que el derecho antidiscriminatorio se ampliase hacia su aplicación a otras formas de la discriminación estructural, como las brechas o la segregación. Más allá de ello, es necesario desarrollar una política pública antidiscriminatoria que pueda garantizar una

acción protectora a la que, hoy por hoy, no pueden acceder las víctimas individualmente ni pueden ofrecer tribunales con poca formación y apoyo.

En el desarrollo de la política antidiscriminatoria, es imprescindible que se coordinen diversos de agentes públicos, a partir de la naciente AESIA, y que se involucren aquellos que ya tienen encargada la lucha a la discriminación. Es el caso del Instituto de la Mujer y para la Igualdad de Oportunidades, la nueva Oficina Antidiscriminación de la Inspección de Trabajo, el Defensor del Pueblo, etc. La expansión y afianzamiento de dicha política dependerá también de la colaboración con actores relevantes de la sociedad civil, como las propias empresas de tecnología, las organizaciones y asociaciones interesadas, los sindicatos o los centros de investigación.

Este desarrollo descansa sobre una premisa de transparencia. Aunque no todo en los algoritmos sea comprensible a la mente humana, es necesario que se sepa cuándo y para qué se utilizan sistemas de IA, y cuál es su impacto. Hace falta también que se asuman las responsabilidades por su uso: que los modelos y los códigos sean objeto de control, y que haya mecanismos de certificación, de seguimiento y de evaluación de impacto. Y, finalmente, es necesario asumir que la IA es una herramienta y no es inevitable: cuando ninguna medida de prevención o mitigación pueda garantizar el derecho a no sufrir discriminación o impedir la vulneración de derechos fundamentales, también se puede limitar, condicionar o prohibir el uso de los sistemas de IA.

Proponemos esta **serie de recomendaciones** para identificar qué es la discriminación algorítmica y hacerle frente.

- **Promover un auténtico enfoque integrado y multidisciplinar de la discriminación.** Las soluciones tecnológicas a los sesgos algorítmicos deben fijar objetivos y criterios concretos y no presentarse como soluciones únicas, omnicomprensivas o que limiten el fenómeno de la discriminación a aquello a lo que ellas pueden dar solución (algunos tipos de sesgos en algunos tipos de sistemas de IA). Fomentar una colaboración que tenga en cuenta la triple dimensión, tecnológica, empírica y ético-política, de la discriminación algorítmica.
- **Analizar la continuidad entre la legislación antidiscriminatoria y la regulación de la IA.** Este análisis es necesario para evitar contradicciones, lagunas y “ventanas de exención” de la aplicación del principio de igualdad y de no discriminación. Incluir la prohibición explícita de la discriminación directa, indirecta e interseccional en la legislación que regula el uso de la IA.
- **Establecer mecanismos de transparencia obligatorios para todos los que utilicen, en el sector público y en el privado, sistemas de IA.** Podrían crearse registros de los sistemas de IA usados, su función y modalidades de aplicación, similares a los registros salariales, de horas extras o de planes de igualdad.
- **Establecer una política pública antidiscriminatoria de aplicación en el contexto de la IA.** Dicha política antidiscriminatoria debería cumplir, al menos, tres funciones: generar conocimiento y evidencia sobre la discriminación resultante del uso de IA, controlar y evaluar el impacto de los sistemas de IA, y apoyar y empoderar a las personas y a las organizaciones para reclamar en casos de discriminación y violación de derechos fundamentales.

- **Dotar a la AESIA de capacidad y recursos para ejercer funciones de control, evaluación y sanción de los casos de discriminación algorítmica.** Establecer funciones de auditoría y certificación de algoritmos de la AESIA, y fomentar la colaboración con organismos públicos y privados de investigación o consultoría. Las auditorías algorítmicas podrían funcionar siguiendo el modelo de las auditorías salariales. El registro y certificación de los sistemas de IA podría ser condición para acceder a fondos o contratos públicos, como ocurre ahora con los planes de igualdad.
- **Coordinar los esfuerzos de diversos actores públicos en la implementación de la política antidiscriminatoria.** Fomentar mecanismos de colaboración entre diferentes autoridades, instituciones y agencias públicas para afrontar la discriminación algorítmica de una forma global, que abarque diversos ámbitos: el de la protección de datos (AEPD), el laboral (Ministerio de Trabajo, Oficina Antidiscriminación de la Inspección de Trabajo), el de la IA (AESIA), y el del género (Instituto de la Mujer).
- **Promover la investigación de la discriminación algorítmica.** Dicha investigación debería ser capaz de generar evidencia sobre el impacto de los sistemas de IA que pueda ser de utilidad a potenciales víctimas, asociaciones y Administraciones para la reclamación y la garantía de los derechos. Promover la investigación basada en la IA para generar conocimiento sobre la discriminación en general y hacer que desarrolle soluciones tecnológicas más éticas que puedan contrarrestar los sistemas comerciales de IA más centrados en la maximización de beneficios.

Referencias

Access Now (2021). The EU should regulate AI on the basis of rights, not risks. [online] Disponible en: <https://www.accessnow.org/eu-regulation-ai-risk-based-approach/> (Consultado: 25-3-2022).

Agencia de los Derechos Fundamentales de la Unión Europea y Consejo de Europa. (2018). #BigData: Discrimination in data-supported decision making. [pdf] Disponible en: <https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making> (Consultado: 9-3-2022).

Agencia de los Derechos Fundamentales de la Unión Europea y Consejo de Europa. (2020). Getting the future right: Artificial Intelligence and Fundamental Rights. [pdf] Disponible en: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf (Consultado: 9-3-2022).

Balayn, A. y Gürses, S. (2021). Beyond Debiasing: Regulating AI and its inequalities. European Digital Rights. [pdf] Disponible en: https://edri.org/wp-content/uploads/2021/09/EDRi_Beyond-Debiasing-Report_Online.pdf (Consultado: 9-3-2022).

Barocas, S. y Selbst, D. (2016). Big Data's Disparate Impact. California Law Review vol. 104. [pdf] Disponible en: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899 (Consultado: 9-3-2022).

Battaglini, M. (2019). Las próximas elecciones votarás a quien mejor te haya manipulado. ¿Crees que eres inmune? Pues no lo eres tanto. LinkedIn. [online] Disponible en: <https://www.linkedin.com/pulse/las-pr%C3%B3ximas-elecciones-votar%C3%A1s-quien-mejor-te-haya-manuela/?originalSubdomain=es> (Consultado: 9-3-2022).

Battaglini, M. (2020). Cómo el Código Postal sirve para la discriminación de precios. LinkedIn. [online] Disponible en: <https://www.linkedin.com/pulse/c%C3%B3mo-el-c%C3%B3digo-postal-sirve-para-la-discriminaci%C3%B3n-de-manuela/> (Consultado: 9-3-2022).

Boulamwini, J. y Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research 81:1–15, 2. [pdf] Disponible en: <https://proceedings.mlr.press/v81/boulamwini18a/boulamwini18a.pdf> (Consultado: 9-3-2022).

Buomso, K. y Park, J. (2020). Transparency and accountability in AI decision support: Explaining and visualizing convolutional neural networks for text information. ScienceDirect. [pdf] Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S0167923620300579> (Consultado: 9-3-2022).

Crenshaw, K. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. University of Chicago Legal Forum. [pdf] Disponible en: <https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1052&context=ucf> (Consultado: 9-3-2022).

Défenseur des droits y Commission Nationale de l'Informatique et des Libertés, CNIL. (2020). Algorithms: preventing automated discrimination. [pdf] Disponible en: https://www.defenseurdesdroits.fr/sites/default/files/atoms/files/836200280_ddd_synthalgos_access.pdf (Consultado: 9-3-2022).

Digital Future Society. (2020). Diseños sensibles a las cuestiones de género en el estado de bienestar digital. [pdf] Disponible en: <https://digitalfuturesociety.com/es/report/exploring-gender-responsive-designs-in-digital-welfare/> Barcelona, España (Consultado: 15-5-2022).

Digital Future Society. (2020). Hacia la igualdad de género en los sistemas digitales de ayudas sociales. [pdf] Disponible en: <https://digitalfuturesociety.com/es/report/hacia-la-igualdad-de-genero-en-el-estado-de-bienestar-digital/> Barcelona, España (Consultado: 15-5-2022).

Digital Future Society. (2022). Gobernanza y algoritmos: riesgos y potencial del uso de la inteligencia artificial en el sector público. [pdf] Disponible en: <https://digitalfuturesociety.com/es/report/governing-algorithms/> Barcelona, España (Consultado: 9-3-2022).

Eguiluz Castañeira, J. A. (2020). Desafíos y retos que plantean las decisiones automatizadas y los perfilados para los derechos fundamentales. Estudios de Deusto, Vol. 68 (2). [pdf] Disponible en: <https://revista-estudios.revistas.deusto.es/article/view/1954> (Consultado: 9-3-2022).

Eubanks, V. (2018). Automating Inequality. How high-tech tools profile, police, and punish the poor. Nueva York, EE. UU., Ed. St Martin's Press.

European Digital Rights, EDRI (2021). Civil society calls for AI red lines in the European Union's Artificial Intelligence proposal. [pdf] Disponible en: <https://edri.org/our-work/civil-society-call-for-ai-red-lines-in-the-european-unions-artificial-intelligence-proposal/> (Consultado: 25-3-2022).

Gerards, J. y Xenidis R. (2020). Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law. Comisión Europea. [pdf] Disponible en: www.equalitylaw.eu/downloads/5361-algorithmic-discrimination-in-europe-pdf-1-975 (Consultado: 9-3-2022).

Gillis, T. y Spiess, J. (2019). Big Data Discrimination. The University of Chicago Law Review, 86: 459–487. [pdf] Disponible en: <https://chicagounbound.uchicago.edu/uclrev/vol86/iss2/4> (Consultado: 25-3-2022).

Grupo independiente de Expertos de Alto Nivel sobre la Inteligencia Artificial. (2020). Lista de Evaluación para una IA Fiable (ALTAI). Comisión Europea. [pdf] Disponible en: <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Consultado: 9-3-2022).

Hacker, P. (2018). Teaching fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law. Common Market Law Review 55:1143-1185. [pdf] Disponible en: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3164973 (Consultado: 9-3-2022).

Hao, K. (2019). Making face recognition less biased doesn't make it less scary. MIT Technology Review. [online] Disponible en: <https://www.technologyreview.com/2019/01/29/137676/making-face-recognition-less-biased-doesnt-make-it-less-scary/> (Consultado: 25-3-2022).

Jiménez, M. (2021). Un buscador para conocer qué algoritmos impactan en tu vida. El País. [online] Disponible en: https://cincodias.elpais.com/cincodias/2021/10/07/companias/1633642034_800454.html (Consultado: 9-3-2022).

- Kleinberg, J. y Sunstein, C. (2019). Discrimination in the Age of Algorithms. *Journal of Legal Analysis*. [pdf] Disponible en: <https://academic.oup.com/jla/article/doi/10.1093/jla/laz001/5476086> (Consultado: 9-3-2022).
- Lagioia, F. y Sartor, G. (2020). Artificial intelligence in the big data era: risks and opportunities. *ResearchGate*. [pdf] Disponible en: https://www.researchgate.net/publication/345373824_Artificial_intelligence_in_the_big_data_era_risks_and_opportunities (Consultado: 9-3-2022).
- Mahler, T. (2021). Between risk management and proportionality: The risk-based approach in the EU's Artificial Intelligence Act Proposal. *Nordic Yearbook of Law and Informatics*. [pdf] Disponible en: <https://ssrn.com/abstract=4001444> (Consultado: 9-3-2022).
- Monclús, J. y Martínez, J. M. (2021). Algoritmos y representación legal de la plantilla: Del secreto empresarial... ¿A la cogobernanza? *Cuatrecasas*. [online] Disponible en: <https://www.cuatrecasas.com/es/spain/articulo/algoritmos-representacion-legal-plantilla-secreto-empresarial-cogobernanza> (Consultado: 9-3-2022).
- National Institute of Standards and Technology, NIST. (2021). Proposal for Identifying and Managing Bias in Artificial Intelligence. NIST Special Publication 1270. [pdf] Disponible en: <https://www.nist.gov/artificial-intelligence/proposal-identifying-and-managing-bias-artificial-intelligence-sp-1270> (Consultado: 9-3-2022).
- Oliva León, R. (2021). ¿Es discriminatorio el algoritmo de Deliveroo que "rankea" a sus riders? *Algoritmo Legal*. [online] Disponible en: <https://www.algoritmolegal.com/tecnologias-disruptivas/es-discriminatorio-el-algoritmo-de-deliveroo-que-rankea-a-sus-riders> (Consultado: 9-3-2022).
- Orwat, C. (2020). Risks of Discrimination through the Use of Algorithms. *ResearchGate*. [pdf] Disponible en: https://www.researchgate.net/publication/344815625_Risks_of_Discrimination_through_the_Use_of_Algorithms (Consultado: 9-3-2022).
- Pandey, A. y Caliskan, A. (2021). Disparate Impact of Artificial Intelligence Bias in Ridehailing Economy's Price Discrimination Algorithms. *arXiv*. [pdf] Disponible en: <https://arxiv.org/abs/2006.04599> (Consultado: 9-3-2022).
- Schiek, D. (2016). Revisiting intersectionality for EU anti-discrimination law in an economic crisis – A critical legal studies perspective. *ResearchGate*. [pdf] Disponible en: https://www.researchgate.net/publication/312241500_Revisiting_intersectionality_for_EU_anti-discrimination_law_in_an_economic_crisis_-_A_critical_legal_studies_perspective (Consultado: 9-3-2022).
- Selbst, A. D. y Powles, J. (2017). Meaningful information and the right to explanation. *International Data Privacy Law*, Volume 7, Issue 4, November 2017, Pages 233–242. [pdf] Disponible en: <https://academic.oup.com/idpl/article/7/4/233/4762325> (Consultado: 9-3-2022).
- Wachter, 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* (7,2). [pdf] Disponible en: <https://academic.oup.com/idpl/article/7/2/76/3860948> (Consultado: 9-3-2022).
- Xenidis, R. y Senden, L. (2020). EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination. *General Principles of EU Law and the EU Digital Order*. [PDF] Disponible en: https://www.researchgate.net/publication/348382658_EU_Non-discrimination_Law_in_the_Era_of_Artificial_Intelligence_Mapping_the_Challenges_of_Algorithmic_Discrimination (Consultado: 9-3-2022).

Zuboff, S. (2020). La era del Capitalismo de la Vigilancia. Madrid, España: Ed. Paidós.

Zuiderveen Borgesius, F. J. (2018). Discrimination, artificial intelligence, and algorithmic decision-making. Consejo de Europa. [pdf] Disponible en: <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73> (Consultado: 9-3-2022).

Zuiderveen Borgesius, F. J. (2020). Strengthening legal protection against discrimination by algorithms and artificial intelligence. The International Journal of Human Rights. [pdf] Disponible en: <https://www.tandfonline.com/doi/pdf/10.1080/13642987.2020.1743976?needAccess=true> (Consultado: 9-3-2022).

Agradecimientos

Autores

Dolores Morondo Taramundi, investigadora principal del Instituto de Derechos Humanos Pedro Arrupe de la Universidad de Deusto y miembro de la Red Europea de Expertos Legales en igualdad y no discriminación. Sus principales áreas de investigación son el derecho antidiscriminatorio, las teorías críticas del derecho, los derechos humanos y la metodología jurídica. Licenciada en Derecho, con especialidad económica y doctorado en Derecho del Instituto Universitario Europeo.

Josu Andoni Eguiluz Castañeira, investigador centrado en los retos jurídicos que las nuevas tecnologías (inteligencia artificial, decisiones automatizadas, metaverso, etc.) plantean para los derechos fundamentales. Graduado en el departamento de Propiedad Intelectual, Nuevas Tecnologías y Protección de Datos de Cuatrecasas. Graduado en Derecho y TIC en la Universidad de Deusto y Doble Máster de Acceso a la Abogacía y Propiedad Intelectual, Nuevas tecnologías y Protección de Datos en Esade Law School.

Equipo de Digital Future Society Think Tank

- **Carina Lopes**, directora de Digital Future Society Think Tank
- **Olivia Blanchard**, investigadora de Digital Future Society Think Tank
- **Tanya Álvarez**, investigadora de Digital Future Society Think Tank

Citas

Este informe se debe citar de la siguiente manera:

- Digital Future Society. (2022). La discriminación algorítmica en España: límites y potencial del marco legal. Barcelona, España.

Datos de contacto

Si desea ponerse en contacto con el equipo de Digital Future Society Think Tank, envíe un correo electrónico a thinktank@digitalfuturesociety.com



**Digital
Future Society**