

Prevenir y controlar la discriminación algorítmica

Autora:

Sánchez Caparrós, Mariana

Cita: RC D 427/2021

Subtítulo:

Sobre regulaciones y otras herramientas disponibles para impugnar la decisión automatizada que discrimina

Encabezado:

La IA plantea numerosos desafíos que el derecho debe abordar con relativa urgencia. Uno de ellos es, sin dudas, la problemática de la discriminación algorítmica, que demanda un esfuerzo inmediato de los Estados para diseñar los marcos regulatorios que atiendan los extremos necesarios para lograr una IA ética, fiable, robusta, equitativa y centrada en el ser humano en el ámbito de LATAM, y otro de los operadores jurídicos (jueces y abogados) para pensar en herramientas que permitan impugnar el sesgo algorítmico que ya está ocurriendo. La doctrina de las categorías sospechosas puede ser una de ellas.

Sumario:

I. Introducción. II. Inteligencia artificial, machine learning, cajas blancas y cajas negras. II.1. ¿De qué hablamos, cuando hablamos de IA? II.2. IA general e IA restringida, Machine Learning, Black & White boxes. III. ¿Algoritmos que discriminan? El problema del sesgo injusto. III.1. De qué hablamos cuando hablamos de sesgo algorítmico. III.2. La problemática del sesgo frente a las cajas negras. IV. Prevenir el sesgo injusto - La regulación como instrumento. IV.1. Principios rectores para una IA ética, robusta, fiable y centrada en las personas IV.2. ¿Quién dicta los marcos éticos? Iniciativas del sector privado y el rol del sector público IV.3. Algunas propuestas regulatorias en el mundo: el Libro Blanco de la UE; la propuesta del Parlamento Europeo y el proyecto de Ley de Responsabilidad Algorítmica en EE.UU. IV.3.1. El Libro Blanco de la UE. IV.3.2. La propuesta del Parlamento Europeo del año 2020. IV.3.3. El proyecto de Ley de Responsabilidad Algorítmica en los EE.UU. V. Impugnar la decisión automatizada que discrimina - Las categorías sospechosas como herramienta. V.1. Las categorías sospechosas en el ámbito latinoamericano y en particular, en el ámbito de la Corte Interamericana de Derechos Humanos. V.2. Categorías sospechosas y potencialmente sospechosas - Relevancia de la distinción. V.3. Las categorías sospechosas para impugnar y controlar el sesgo algorítmico. VI. Conclusiones. VII. Bibliografía.

Prevenir y controlar la discriminación algorítmica

I. Introducción

Cuando la toma de decisiones es sesgada puede tener un impacto negativo en las personas, vulnerando su derecho a la igualdad y a la no discriminación.

Por este motivo, a lo largo de la historia, hemos puesto un notable esfuerzo en trabajar en los distintos ámbitos en los que se adoptan decisiones que pueden afectarnos, como ser el comercio, la justicia, la administración y la legislación, entre otros, generando herramientas dirigidas a mitigar los sesgos injustos, promover la equidad y resguardar el derecho a la igualdad de las personas.

En la actualidad, para la toma de decisiones, la explosión en el volumen y disponibilidad de datos, así como la mayor accesibilidad a algoritmos de *machine learning* que cada vez son más sofisticados, brinda poderosas herramientas para identificar dónde está ocurriendo el sesgo -entendido como discriminación-, y medir si los instrumentos empleados para combatirlo están siendo efectivos^[1].

No obstante, esos mismos datos tienen un lado oscuro, pues como veremos seguidamente, y aquí un "alerta spoiler", cuando se incorporan en sistemas informáticos que utilizan IA, puede suceder -y de hecho sucede- que las decisiones que se adopten con esos sistemas terminen por amplificar y profundizar ciertos sesgos que vienen junto a los datos.

En otras palabras, cuando se trata de decisiones adoptadas con sistemas que utilizan inteligencia artificial -IA-, la subjetividad que históricamente se ha presentado en la decisión humana y que hemos tratado de neutralizar con distintas herramientas para que no se refleje en un tratamiento discriminatorio hacia determinadas personas o grupos, puede tener efectos mucho más amplios y afectar y discriminar a muchas más personas, sin que existan los mecanismos de control social y de autolimitación que, en cambio, si están presentes frente al comportamiento humano[2].

Los puntajes crediticios, por ejemplo, se han utilizado durante décadas para evaluar la solvencia crediticia del consumidor. No obstante, su alcance es mucho mayor ahora que los sistemas creados para valorarlos funcionan con algoritmos de *machine learning*, pues no sólo consideran muchos más datos para hacer la evaluación -tanto en volumen como en tipo-, sino que tienen mayor extensión, afectando cada vez más la posibilidad de comprar un coche, alquilar un apartamento o conseguir un trabajo de tiempo completo[3].

En estos casos, se viene observando que el rápido crecimiento y la adopción de sistemas automatizados para la toma de decisiones, ha creado para las personas de bajos ingresos una red oculta de trampas entrelazadas que comienza a excluirlos de algunos servicios considerados básicos[4].

Del mismo modo, se ha detectado que existen sistemas informáticos que utilizan IA para hacer sugerencias de compras a posibles consumidores que presentan sesgos en la consideración de la diversidad sexual, lo que se ha visto reflejado en el hecho de que estos sistemas, por ejemplo, daban por sentado que todos los clientes del sitio web de una determinada tienda eran heterosexuales, y en función de ello efectuaban sugerencias de presentes para San Valentín para personas del género opuesto al del usuario, privando de ofertas adecuadas a quienes tienen pareja del mismo sexo[5].

Ha sucedido también que una médica, luego de que cambiaran el modo de acceso a los cambiadores de su gimnasio, migrándolo hacia un sistema informático gestionado por un tercero, no pudo ingresar al cambiador de mujeres ya que su tarjeta de acceso era rechazada por un algoritmo en el que el título "Doctor" había sido codificado como masculino[6].

Y pudo observarse como aplicaciones de reconocimiento de imágenes como Amazon Rekognition, no etiquetan del mismo modo -como cabría esperar que lo hicieran- los elementos de una fotografía cuando aparecen hombres y cuando aparecen mujeres, y así reconocen un taladro como herramienta en una fotografía en la que hay un hombre y en cambio, cuando en la misma fotografía aparece una mujer, esa herramienta no es reconocida por el algoritmo[7].

Los ejemplos brindados, que no son los únicos que existen pues hay muchísimos otros[8], son suficientes para demostrar que debemos comenzar a poner el foco en esta problemática y abordar su estudio desde un punto de partida en el que todos habremos de coincidir, y es que la tecnología debe siempre ser utilizada para las personas y no contra ellas, y por ello los sistemas de IA deben diseñarse, elaborarse y desplegarse de un modo que los haga compatibles con los derechos humanos[9].

En consecuencia, las características distintivas de la IA siempre deben ser escrutadas desde la perspectiva de los derechos humanos[10], lo que nos obliga a indagar, entre muchos otros temas relevantes cuando de IA y Derecho se trata, acerca de cómo los datos influyen en el comportamiento de los sistemas de IA, para así poder diseñar herramientas adecuadas para prevenir, mitigar y combatir el sesgo algorítmico.

II. Inteligencia artificial, machine learning, cajas blancas y cajas negras

II.1. ¿De qué hablamos, cuando hablamos de IA?

Es menester advertir que todavía no existe consenso en torno a cómo definir la IA, y por ello, según la noticia, texto o autor que estemos leyendo, encontraremos diversas acepciones que se han construido en razón de distintos objetivos perseguidos por el sistema sobre el que se habla, o por el estado de la técnica al momento de describirla[11].

No obstante, recientemente, en el ámbito de la Comisión Europea, el Grupo de Expertos de Alto Nivel en IA ha consensado una definición que se considera lo suficientemente flexible como para adaptarse al progreso continuo que lleva esta técnica y que en paralelo conserve un adecuado nivel de precisión, que brinde la seguridad jurídica que debe tener cualquier definición de un concepto que se adopta en el orden legal[12].

Entonces, ¿qué son los sistemas de IA?

El Grupo de Expertos de Alto Nivel en IA los ha definido como "...sistemas de software (y posiblemente también hardware) diseñados por humanos que, dado un objetivo complejo, actúan en la dimensión física o digital, percibiendo su entorno a través de la adquisición de datos, interpretando datos estructurados o no estructurados recopilados, razonando sobre el conocimiento, o procesando la información, derivada de esos datos, y decidiendo cuál es la mejor acción a tomar para lograr un objetivo dado.

Los sistemas pueden usar reglas simbólicas o aprender un modelo numérico, y también pueden adaptar su comportamiento analizando cómo el entorno se ve afectado por sus acciones anteriores.

Como disciplina científica la IA incluye varios enfoques y técnicas, como el aprendizaje automático (del cual el aprendizaje profundo y el aprendizaje reforzado son ejemplos específicos), el razonamiento automático (que incluye planificación, programación, representación y razonamiento del conocimiento, búsqueda y optimización) y robótica (que incluye control, percepción, sensores y actuadores, así como la integración de todas las demás técnicas en sistemas ciberfísicos"[13].

Cualquiera sea la definición que se escoja, lo cierto es que en todos los casos, en el corazón de la IA están los algoritmos, que son códigos informáticos diseñados y escritos por seres humanos que logran ejecutar instrucciones a partir de traducir los datos de los que se nutren en conclusiones, información o productos¹⁴; que pueden involucrar diferentes modos de razonamiento epistémico o práctico, y más adelante, pueden auto mejorarse mediante el desarrollo de nuevos heurísticos, modificar su información interna e incluso generar algoritmos nuevos[14].

En otras palabras, las diversas técnicas de IA "...se basan en detectar y reconocer patrones de información en los datos...", lo que "...se logra a partir de combinar ordenadores, internet, algoritmos y lenguajes de programación para resolver problemas o tomar decisiones que antes solo podían ser realizadas por nuestras capacidades cognitivas..."[15].

II.2. IA general e IA restringida, Machine Learning, Black & White boxes

Los sistemas de IA que conocemos hasta el momento pertenecen al ámbito de la IA restringida o débil, que es aquella que permite a los sistemas informáticos ejecutar tareas programadas y brindar soluciones en dominios particulares del conocimiento humano[16].

En otras palabras, la IA restringida comprende un conjunto de técnicas informáticas dirigidas a lograr un resultado específico en áreas determinadas que antes sólo podíamos lograr los seres humanos[17].

Como contrapartida, la IA general da cuenta de sistemas que aún no existen, y alude a una IA capaz de reproducir nuestra habilidad para gestionar varios campos de conocimiento a la vez, así como nuestro sentido común. Es decir, que simule el comportamiento humano[18].

El *machine learning* o aprendizaje automatizado, por su parte, es un género dentro de la IA débil, que comprende un conjunto de técnicas y métodos que permiten a los algoritmos extraer correlaciones de los datos con mínima o ninguna supervisión, y que agrupa varias especies, algunas de las cuales son de caja blanca y otras de caja

negra[19].

Los sistemas de caja blanca o *white boxes* son sistemas de IA en los que la explicabilidad aparece como una propiedad del sistema que permite que pueda proporcionar una forma de explicación de sus acciones[20] que puede ser comprendida por expertos humanos.

Como contrapartida, la noción de IA de caja negra o *black box* se vincula con aquellos escenarios en los que los sistemas de IA no nos permiten rastrear el motivo de ciertas decisiones adoptadas con ellos, sea de modo total o parcial.

En otros términos, la idea de caja negra refiere a enfoques de aprendizaje automático en los que los pasos a través de los cuales el sistema de IA ha llegado a una decisión, son inteligibles para los expertos humanos, incluidos quienes han diseñado el sistema[21].

Esta característica de los sistemas de IA de caja negra, hace se presenten como un importante obstáculo para la efectividad del principio de explicabilidad[22] cuando se trata de escrutar los resultados de su aplicación, y conlleva la necesidad de prestarles particular atención desde el derecho, sobre todo cuando estos sistemas se emplean para tomar decisiones que afectan los derechos de las personas.

Un típico ejemplo de IA de caja negra lo encontramos en los sistemas de IA basados en redes neuronales complejas, también conocido como *deep learning* o aprendizaje profundo.

Este tipo de sistema, si bien tiene la ventaja de ser más eficiente para lograr resultados porque puede reconocer patrones en cantidades masivas de datos y procesar información no estructurada (v.gr. imágenes o audio), también representa un enorme desafío para los derechos humano, ya que para que la IA sea fiable es esencial comprender por qué un sistema se ha comportado de una determinada manera y por qué ha ofrecido una respuesta específica[23], y éste es justamente el talón de Aquiles de una técnica que, por su propia naturaleza, conlleva que no sea posible explicar los millones de correlaciones "...que se procesan en las capas ocultas de la red (...) en el sentido de que se pueda ofrecer una explicación detallada de lo que ocurrió allí..."[24].

III. ¿Algoritmos que discriminan? El problema del sesgo injusto

III.1. De qué hablamos cuando hablamos de sesgo algorítmico

Históricamente, los conceptos de "sesgo" -*bias*- y "justicia" -*fairness*-, han sido empleados por los profesionales de la informática para describir diferentes problemas éticos ligados a la operación y los resultados de decisiones algorítmicas. Así, la noción de "sesgo algorítmico" se aparece más amplia que la de discriminación que operamos en el ámbito jurídico, ya que no sólo comprende los errores injustos, sino toda clase de errores sistemáticos, tanto de tipo estadístico, estructural, cognitivo o social[25].

Ahora bien, invocado en el contexto de lo justo, el "sesgo algorítmico" comprende un tipo particular de error: aquel que coloca a los grupos privilegiados en una situación de ventaja sistemática y a los grupos no privilegiados en una situación desventaja sistemática[26]. Aquí se asimila más a la noción de discriminación, entendida como el trato desfavorable que se otorga a un individuo o a un grupo de individuos con relación a otros[27].

De cualquier manera, el término "sesgo algorítmico" es más amplio en cuanto a sus alcances que el término legal "discriminación algorítmica". El primero, alcanza cualquier tipo de desventaja que pueda ser vista como moral o éticamente reprochable. El segundo, en cambio, comprende el trato desfavorable formulado sobre la base de categorías específicamente protegidas por las normas, sea explícitamente o implícitamente[28].

A los fines de este trabajo, el concepto de sesgo se refiere particularmente al "sesgo injusto", es decir, a aquella inclinación como prejuicio o estereotipo que perjudica a una persona o grupos de personas, sea directa o indirectamente, dando lugar a resultados discriminatorios[29].

Los sistemas algorítmicos, sabemos, se nutren de los datos, son capaces de analizarlos en grandes volúmenes rápidamente y con ello hacer que los programas de IA puedan adoptar decisiones que antes eran de dominio de las personas que las ejecutaban sin herramientas informáticas[30].

Sin embargo, esos algoritmos pueden reforzar los sesgos presentes en su diseñador, que inadvertidamente los introduce al entrenar el sistema; o hacer lo propio con aquellos que vienen con los conjuntos de datos que utilizan los sistemas de IA, sea con fines de información o para su funcionamiento, ya que esos datos pueden presentar sesgos históricos inadvertidos, cuyo mantenimiento podría dar lugar a prejuicios y discriminación contra determinados grupos o personas, agravando los estereotipos ya existentes e incrementando su marginación[31].

Dicho de otro modo, diversos factores pueden colaborar en la incorporación del sesgo injusto en los sistemas de IA, aumentando su potencial discriminatorio. Entre ellos podemos identificar "...la forma en que se diseñan los sistemas, las decisiones sobre el origen y el alcance de los conjuntos de datos con que se entrenan, los sesgos sociales y culturales que los creadores de aplicaciones pueden incorporar en esos conjuntos de datos, los modelos mismos de IA y la forma en que los productos del modelo de IA se ejecutan en la práctica..."[32].

Partiendo de esa idea, el Prof. Amunátegui Perelló ha identificado tres tipos de sesgos que un sistema de IA puede adquirir con relación a los datos, estos son: a) el sesgo de interacción; b) el sesgo latente; y c) el sesgo de selección[33].

El *sesgo de interacción* se presenta cuando el propio usuario o programador introduce inadvertidamente un sesgo en el modelo por el modo en que interactúa con él, por ejemplo, al definir la función de éxito del modelo o su objeto[34].

Así, para un sistema de IA para la selección de postulantes a una universidad, si el programador define una función de entrada favorable que aplica sólo a miembros de ciertos grupos, como podría ser el provenir de ciertas instituciones educativas consideradas de nivel elevado[35], estaremos frente a un claro sesgo de interacción.

El *sesgo latente*, por su parte, aparece cuando el sistema de IA efectúa correlaciones inapropiadas entre los datos, dando lugar a falsos nexos entre distintos puntos de datos[36].

Así, el sesgo latente podría presentarse, por ejemplo, en el caso de que un gerente de recursos humanos de una firma no hubiera contratado durante su gestión a personas pertenecientes a un determinado grupo étnico y coincidiera que esas personas suelen vivir en determinados vecindarios. En ese caso, el *set* de entrenamiento basado en decisiones anteriores de ese mismo gerente le "enseñaría" al sistema a no seleccionar personas que habiten esos vecindarios al correlacionar ambos puntos de datos, automatizando una tarea predictiva que profundizará o consolidará el rechazo de las solicitudes que provengan de ese grupo de individuos[37].

Es importante tener presente, llegados a este punto, que correlación y causalidad en la materia que tratamos no son conceptos que puedan tomar como equivalentes.

La causalidad, recordemos, examina la relación que existe entre dos variables para tratar de identificarla, y en cambio, la correlación da cuenta de que puede haber un patrón en los datos que lleva a que se muevan de manera conjunta. Por ello, que exista correlación entre variables, no siempre significa que haya una relación de causalidad entre ellas, es decir, que una variable sea la que causa la otra.

Veámoslo con un ejemplo: yo puedo observar que cuando canta el gallo de mi vecina amanece -hay aquí una correlación-. No obstante, eso no significa que el canto del gallo de mi vecina sea lo que provoca -o causa- el amanecer. Es decir, no hay relación de causalidad entre el canto del gallo y el amanecer, más allá de que pueda correlacionar ambos hechos porque muy a menudo se presentan juntos.

En la materia que nos ocupa, se ha observado que el aprendizaje automático es muy útil para encontrar patrones o correlaciones estadísticas entre los diversos datos con los que trabaja. No obstante, esto no quiere decir que comprenda la causalidad, es decir, la relación causa-efecto que puede haber -o no- entre esos datos.

Finalmente, el *sesgo de selección* se produce cuando el set de datos no es lo suficientemente representativo de la diversidad existente en el medio social[38].

Un excelente ejemplo provisto por Amunátegui Perelló grafica este sesgo: "...si un algoritmo se entrena para determinar elementos que hagan predecir habilidades en una población para ser buenos jugadores de fútbol con los datos médicos de todos los jugadores de la primera división argentina, probablemente dicho algoritmo sea inútil para realizar esa predicción en Japón, debido a su baja representatividad de población asiática..."[39].

Los ejemplos dados hasta aquí, ponen en evidencia que aunque suele presumirse que los sistemas de IA son neutros y deciden objetivamente, lo cierto es que se entrenan con datos, y si estos datos incluyen sesgos injustos, estos sistemas van a amplificar estructuralmente esos prejuicios provenientes de los datos, consolidando y expandiendo la discriminación[40], dando lugar a un círculo vicioso que terminará por excluir definitivamente a muchas personas del acceso a distintos bienes y servicios, sobre todo a aquellas que pertenecen a grupos vulnerables e históricamente desfavorecidos, reiniciando una problemática que desde nuestra disciplina venimos combatiendo hace largo tiempo.

III.2. La problemática del sesgo frente a las cajas negras

La problemática del sesgo injusto es un campo de investigación en curso para el derecho que plantea diversos desafíos según la técnica de *machine learning* que se utilice para tomar decisiones.

En efecto, es esta una problemática que se volverá especialmente interesante frente a sistemas que empleen técnicas de IA de caja negra, porque en esos casos la discriminación va a ser mucho más difícil o a veces prácticamente imposible de escrutar, ya que es técnicamente impracticable trazar en un cien por cien cómo estos sistemas han llegado a la decisión adoptada o sugerida[41].

Así, cuando de cajas negras se trata, se torna prácticamente imposible para los profesionales del derecho, entre ellos abogados y jueces, entender el razonamiento detrás del resultado arrojado por el sistema, complicando no sólo el motivar la decisión adoptada con o en base al sistema, sino también la posibilidad de cuestionarla adecuadamente y de abordar esa impugnación[42].

En contraposición con ello, si estamos frente a técnicas de caja blanca, siempre que no se opongan motivos de propiedad intelectual para acceder al código fuente, la decisión será auditable -idealmente debería ser públicamente auditable-, lo que dará lugar a un mayor control y a que las decisiones basadas en algoritmos puedan ser más justas[43].

Los desafíos que nos plantean las diversas técnicas de *machine learning* en cuanto a su explicabilidad, ponen en evidencia que basarse y confiar demasiado en decisiones automatizadas, sin reconocer la problemática de los sesgos, puede impedir que se realice un examen crítico de los resultados de la IA y obstaculizar el acceso de las personas a recursos contra decisiones adversas tomadas por estos sistemas inteligentes[44].

IV. Prevenir el sesgo injusto - La regulación como instrumento

Ahora que sabemos que el problema de los sesgos existe, que los algoritmos no son neutrales y que debemos atender esta problemática, el paso que sigue es pensar en qué herramientas podemos desarrollar desde nuestra disciplina a los fines de prevenir y mitigar el sesgo algorítmico.

En otras palabras, debemos pensar en herramientas que prevengan que los actores sociales que utilizan sistemas de IA para tomar decisiones o para dar soporte a sus decisiones, sean públicos o privados, discriminen a los consumidores o potenciales consumidores, a los ciudadanos y al público en general cuando interactúa con ellos.

IV.1. Principios rectores para una IA ética, robusta, fiable y centrada en las personas

Encuadrar la IA en el marco del Estado constitucional de derecho presupone que el diseño, desarrollo y uso de

los sistemas de IA respete la dignidad humana y los derechos humanos[45].

Partiendo de allí, y considerando que estamos frente a una tecnología en pleno desarrollo, que además es muy dinámica, pues cambia constantemente, el marco jurídico que se proponga debería basarse en el respeto de los derechos y en el establecimiento de una serie de principios básicos y fundamentales que sienten las bases para una IA ética, robusta, fiable y centrada en el ser humano.

Estos principios son especialmente relevantes durante la fase de diseño del sistema, en tanto es en este momento en el que se configuran y programan sus funcionalidades y objetivos[46].

Al respecto, se ha resaltado que es durante la fase inicial, de desarrollo del sistema, que pueden abordarse de manera más efectiva algunos de los principales riesgos que plantea la IA, y es por ello que frecuentemente se habla de mecanismos de *value by design*, *security by design*, y para el caso, *ethics by design* o ética por diseño, que figura la idea de que estos principios éticos se incorporen expresamente en esta primera etapa crítica, que es la del diseño y programación de los sistemas que utilizan técnicas de IA. Ello sin perjuicio de que también alcancen las fases de desarrollo, implementación y adopción del sistema[47].

En el ámbito de la Comunidad Europea, inspirados en los derechos fundamentales consagrados en diversos Tratados de Derechos Humanos, se han identificado cuatro principios imperativos éticos que deberían observarse en todo momento y fundamentalmente, entiendo, en la etapa de diseño del sistema:

1. *Respeto de la autonomía humana*. Los sistemas de IA deben garantizar el respeto de la libertad y la autonomía humana, evitando subordinar, manipular, condicionar o dirigir injustificadamente a los seres humanos. Por ello, deberían operar como un complemento que potencie sus capacidades y aptitudes cognitivas, sociales y culturales, lo que exige una supervisión humana constante[48].

2. *Prevención del daño*. los sistemas de IA no deberían provocar daños o agravar los existentes, ni perjudicar de cualquier otro modo a los seres humanos. Deben ser seguros y robustos desde el punto de vista técnico, garantizándose que no pueden ser objeto de un uso malintencionado. Se debe prestar atención a los efectos adversos que pueden venir dados por la asimetría de las relaciones entre grupos de personas (v.gr. trabajadores y empleadores, consumidores y empresarios)[49].

3. *Equidad y no discriminación algorítmica*. El desarrollo, despliegue y utilización de sistemas de IA debe ser equitativo, garantizando una distribución justa e igualitaria de los costos y beneficios, y asegurar que las personas no sufran discriminación injusta. Las personas deben poder oponerse a las decisiones adoptadas por los sistemas de IA y las personas que operan con ellos, así como tratar de obtener compensaciones adecuadas, identificar a los responsables y solicitar explicaciones[50].

4. *Explicabilidad*. Los procesos de desarrollo, despliegue y uso de IA deben ser transparentes; debe comunicarse abiertamente la capacidad y finalidad de los sistemas de IA, y las decisiones adoptadas por ellos o por las personas que los emplean deben ser explicables, en la medida de lo posible[51].

La equidad y la explicabilidad, entiendo, son dos principios clave cuando se trata de prevenir, mitigar y cuestionar el sesgo presente en los algoritmos de los sistemas de IA.

Podríamos sumar a ellos, la necesidad de garantizar la inclusión y la diversidad a lo largo de todo el ciclo de vida de los sistemas de IA, y la igualdad de acceso mediante procesos de diseño inclusivo[52].

Además, siempre que sea posible, los sesgos identificables y discriminatorios deberían eliminarse en la fase de recopilación de datos, y de no ser posible esto último, debería velarse por una supervisión constante que permita analizar y abordar el propósito, las restricciones, los requisitos y las decisiones del sistema de un modo claro y transparente[53].

A estos fines, la contratación de personas procedentes de diversos contextos, culturas y disciplinas puede ser clave en orden a garantizar la diversidad de opiniones, por lo que debería fomentarse[54].

IV.2. ¿Quién dicta los marcos éticos? Iniciativas del sector privado y el rol del sector público

Existen varias iniciativas del sector privado para regular la IA. La mayoría, como sucede con el consorcio conformado por Google, Amazon, Facebook y Microsoft, al que luego se sumó Apple[55], se centran en la consagración de una serie de principios éticos que tienden a que la IA se diseñe, desarrolle e implemente en beneficio de la sociedad.

No obstante, lo cierto es que aun cuando esas iniciativas son bien recibidas, no dejan de estar producidas por esas mismas corporaciones que se benefician con la explotación de esta tecnología, lo que las coloca en un claro lugar de conflicto de interés al momento de diagramar la regulación, pero sobre todo al momento de aplicarla.

La circunstancia antes dicha es determinante para que los Estados no puedan desentenderse de su deber de garantizar el cumplimiento de los derechos y garantías que las Constituciones de sus países y los instrumentos internacionales sobre Derechos Humanos a los que han adherido reconocen a sus habitantes, y demanda que deban intervenir en la regulación de la IA, generando un marco jurídico que permita supervisar esta tecnología, tal y como sucede con muchas otras tecnologías complejas, que sea imperativo, a la vez que prudente y cuidadosamente limitado[56].

Ello pues dejar esta regulación en manos del sector privado, no sólo que deja a la libre elección de las corporaciones el atenerse o no a las reglas creadas, sino que además, podría conducir a la reproducción de marcos diversos, uno por cada empresa o individuo, lo que sería totalmente caótico, a la vez que peligroso[57].

En ese marco, la IA requiere que los Estados, respaldados por la legitimación que la democracia brinda a sus instituciones, tomen parte en el establecimiento de un marco regulatorio dirigido a asegurar que esta tecnología sea segura y contribuya al bienestar general de la sociedad, y especialmente, a evitar los efectos no deseados que su uso creciente puede provocar, entre los que sin dudas se encuentra el de la discriminación algorítmica, que nos ocupa en este trabajo, entre muchos otros (v.gr. *deep fakes*, responsabilidad algorítmica, privacidad de datos).

IV.3. Algunas propuestas regulatorias en el mundo: el Libro Blanco de la UE; la propuesta del Parlamento Europeo y el proyecto de Ley de Responsabilidad Algorítmica en EE.UU.

IV.3.1. El Libro Blanco de la UE

El Libro Blanco sobre IA, publicado por la Comisión Europea de DDHH en febrero del año 2020, propicia que los marcos regulatorios que se dicten, por un lado, promuevan la adopción de la IA, y por otro, aborden adecuadamente los riesgos asociados con ciertos usos de esta nueva tecnología, partiendo de la base de que las características particulares de algunas técnicas de aprendizaje automático, como su opacidad, complejidad, imprevisibilidad y un comportamiento parcialmente autónomo, pueden dificultar la comprobación del cumplimiento de la regulación protectora de los derechos fundamentales e impedir su efectivo cumplimiento[58].

Para ello, propicia que el marco regulador a dictar se centre en cómo minimizar los riesgos que suponen los sistemas con IA para los derechos fundamentales, especialmente la protección de datos personales, la privacidad y en lo que a este trabajo interesa, la no discriminación[59].

En otras palabras, el marco regulador debe centrarse en cómo minimizar los distintos riesgos de sufrir daños, especialmente los más significativos[60], para lo cual podría partirse de un enfoque basado en el mayor o menor riesgo que entrañan las diversas aplicaciones de IA[61].

En cualquier caso, se señala en el Libro Blanco, si se apela al diseño de normas que regulen de forma directa las aplicaciones de IA, se debe ser cuidadoso y considerar los siguientes aspectos relevantes:

1. La regulación debe ser lo suficientemente flexible como para lograr el fin que persigue, sin desalentar el empleo y desarrollo de esta tecnología[62].

2. Debe ser diseñada de un modo transdisciplinario, con apoyo de profesionales que conozcan sobre IA, *machine learning*, aprendizaje profundo, *data science* y propiedad intelectual, de modo tal que surja una regulación que sea concreta y que sea efectivamente aplicable[63].

3. Se debe considerar que esta es una tecnología en constante evolución, lo que lleva a que normas demasiado precisas o específicas puedan caer en desuso en unos pocos años[64].

4. La regulación debe ser eficaz para alcanzar los objetivos propuestos, sin ser excesivamente prescriptiva, ya que ello podría generar una carga desproporcionada, en especial para las pequeñas y medianas industrias[65].

Además, como se señalaba, en el Libro Blanco se sugiere regular a partir de considerar el mayor o menor riesgo que conlleven las aplicaciones de IA, para lo cual deberán considerarse, por un lado, los derechos que estén en juego, y por otro, el sector y el uso para el que esas aplicaciones fueron previstas[66].

En ese esquema, una aplicación de IA, sin dudas, será de riesgo elevado cuando:

1. Se la utilice en un sector respecto del cual, atendiendo a las actividades que desarrolla, es previsible que existan riesgos significativos (v.gr. sector público, salud, seguridad, energía) -riesgo sector-, y;

2. Se la emplee de una forma que pueda importar riesgos significativos (v.gr. una cosa es un sistema de turnos de un hospital público y otro la gestión del historial médico del individuo) -riesgo de uso-[67].

Ello no excluye que considerando exclusivamente los derechos en juego en cada caso (v.gr. derechos fundamentales, derechos de consumidores y usuarios, derechos de niños, niñas y adolescentes), la aplicación de IA pueda considerarse igualmente de riesgo elevado en función de su uso, independientemente del sector en el que se la utilice[68].

En lo que es de relevancia en materia de prevención y mitigación del sesgo algorítmico, el marco regulatorio según el Libro Blanco debería establecer requisitos con relación a:

1. Los datos de entrenamiento del sistema, procurando que el uso posterior del sistema de IA no genere resultados que conlleven una discriminación ilícita, para lo cual debería establecerse la obligación de utilizar conjuntos de datos de entrenamiento suficientemente representativos, garantizando que todas las dimensiones de género, etnicidad y otros posibles motivos de discriminación queden reflejadas en esos datos[69].

2. La conservación de registros y datos que se utilizaron para programar los algoritmos, para entrenar y testear sistemas de riesgo elevado; de modo tal de poder conocer las características del conjunto de datos y el modo en que se los escogió, o incluso el propio conjunto de datos en ciertos supuestos; o bien las metodologías de programación, entrenamiento, validación y testeo de los algoritmos[70].

3. El acceso a la información con respecto a las posibilidades y limitaciones de la IA, los objetivos para los que fue programada, la aplicación y el nivel de exactitud que se estima conseguir en el logro de ese objetivo, tanto para los implementadores, como para las autoridades competentes y partes afectadas, según el caso[71].

4. El conocimiento del usuario acerca de cuándo está siendo sujeto de una decisión automatizada y cuando, en cambio, existe interacción con, o intervención de, un ser humano[72].

5. La reproducibilidad de los resultados de la aplicación del sistema de IA utilizado[73].

6. La supervisión humana constante, que garantice una IA fiable y antropocéntrica, que no provoque efectos adversos; y que cuando se trata de sistemas de riesgo elevado pueda garantizarse obligatoriamente el control del resultado del sistema previo a que se haga efectivo (v.gr. para el otorgamiento de una prestación de seguridad social) o *a posteriori* (v.gr. para el control de la denegatoria de un crédito bancario), según el caso; en la intervención humana en tiempo real, con posibilidad de desactivar el sistema si se vuelve inseguro (v.gr. en un

auto autónomo, poder apagarlo); en la fase de diseño mediante el establecimiento de restricciones operativas al sistema de IA (v.gr. auto autónomo que se desactiva y vuelve a conductor manual cuando los sensores advierten condiciones de manejo autónomo no seguras)[74].

En todos los casos, el destinatario de la regulación, es decir, quien deberá cumplir con los recaudos que se determinen en los marcos regulatorios, debería ser quien esté en mejor posición de abordar cada posible riesgo considerando las diversas participaciones que pueden tener figuras como el desarrollador, el implementador, el productor, distribuidor o importador, el proveedor de servicios, los usuarios profesionales o particulares[75].

IV.3.2. La propuesta del Parlamento Europeo del año 2020

El 20 de octubre del año 2020 el Parlamento Europeo aprobó tres propuestas de regulación sobre inteligencia artificial, referidas a tres áreas específicas: las normas éticas que deben regir la IA; las reglas sobre la responsabilidad derivada del uso de sistemas de IA; y las reglas sobre derechos de propiedad intelectual vinculadas con el uso de sistemas de IA[76].

La propuesta para un marco ético para la inteligencia artificial, que fue aprobada por 559 votos a favor, 44 en contra y 88 abstenciones, insta a la Comisión Europea a diseñar un nuevo marco jurídico que desarrolle los principios éticos y obligaciones jurídicas vinculados con el desarrollo, despliegue y empleo de la IA, la robótica y tecnologías relacionadas en el ámbito de la UE[77].

Destaca el Parlamento Europeo en su propuesta que sin perjuicio de la legislación sectorial, resulta necesario contar con un marco regulatorio armónico basado en el Derecho de la UE, la Carta Europea y el Derecho internacional de Derechos Humanos, con la finalidad de establecer normas que sean iguales para toda la UE y protejan eficazmente sus valores; que persigan el respeto de la dignidad humana, la autonomía y la autodeterminación de las personas; la prevención del daño y la eliminación de los sesgos y la discriminación, entre otros objetivos destacados[78].

La inteligencia artificial, la robótica y las tecnologías conexas deben adaptarse a las necesidades humanas. Su desarrollo, despliegue y uso deben estar siempre al servicio de las personas y nunca al revés, y deben tener por objeto aumentar el bienestar y la libertad individual, preservar la paz, prevenir los conflictos y reforzar la seguridad internacional, maximizando al mismo tiempo los beneficios ofrecidos y evitando y reduciendo los riesgos[79].

Puntualmente, respecto de la temática atinente a los sesgos, reconoce el Parlamento Europeo que la IA, según su desarrollo y los usos que se le asignen, puede crear y reforzar sesgos, y también discriminar como consecuencia de los sesgos que son inherentes a los conjuntos de datos empleados para entrenar los sistemas; lo que demanda requerir a la Comisión Europea y a los Estados miembros que adopten las medidas pertinentes a los fines de evitarlos y garantizar la plena protección de los derechos fundamentales[80].

En línea con ello, manifiesta su preocupación por el peligro que pueden significar los sesgos y la discriminación en el desarrollo y uso de la IA, y sostiene que los valores éticos de la equidad, la exactitud, la confidencialidad y la transparencia deben ser la base de esta tecnología, lo que implica en este contexto que sus operaciones deban concebirse de un modo tal que garantice la igualdad de trato y la no discriminación[81].

En este marco, señala, adquiere especial relevancia la calidad de los conjuntos de datos utilizados por los sistemas de IA, robótica y tecnologías conexas, según su contexto, "...en particular en lo que respecta a la representatividad de los datos de entrenamiento, así como sobre la eliminación de sesgos en los conjuntos de datos, sobre los algoritmos utilizados y sobre las normas en cuanto a datos y agregación (...) estos conjuntos de datos deben ser auditados por las autoridades nacionales de control siempre que se les solicite para asegurarse de su conformidad con los principios citados anteriormente..."[82].

Finalmente, remarca la relevancia de garantizar que las personas cuenten con recursos efectivos, y que los Estados velen por el establecimiento de procedimientos y mecanismos de revisión accesibles, asequibles, independientes y eficaces para garantizar la revisión imparcial, efectuada por seres humanos, de las denuncias

de vulneración de los derechos de ciudadanos y consumidores, realizada por sistemas algorítmicos, independientemente de si tienen su origen en agentes del sector público o del sector privado[83].

La propuesta de reglamento sobre principios éticos para el desarrollo, el despliegue y el uso de la inteligencia artificial, la robótica y las tecnologías conexas, incluye un artículo específico dirigido a regular lo atinente a la cuestión del sesgo y la discriminación.

Para comenzar, la propuesta regulatoria parte por definir al "sesgo" como "toda percepción personal o social prejuiciosa de una persona o de un grupo de personas sobre la base de sus características personales"[84].

Y en concreto, en el art. 9 prevé que los programas informáticos, los algoritmos o los datos utilizados o producidos por la IA, la robótica y las tecnologías conexas desarrolladas, implementadas o utilizadas en la UE, deberán ser imparciales y no podrán discriminar por motivos de raza, sexo, orientación sexual, embarazo, discapacidad, características físicas o genéticas, edad, minoría nacional, origen étnico o social, lengua, religión o creencias, opiniones políticas o participación cívica, nacionalidad, estado civil o económico, educación o antecedentes penales[85].

En todo caso, prevé propuesta, sólo podrá justificarse el trato diferenciado entre personas o grupos de personas cuando exista una finalidad objetiva, razonable, legítima, proporcionada y necesaria, en la medida en que no exista otra alternativa que interfiera menos con el principio de igualdad de trato[86].

IV.3.3. El proyecto de Ley de Responsabilidad Algorítmica en los EE.UU.

En el año 2019 se presentó en el Congreso de los EE.UU. un proyecto de ley de responsabilidad algorítmica[87], con el objeto de exigir a las empresas que evalúen periódicamente los sistemas de IA que utilizan, a fin de determinar la precisión, justicia, el sesgo y la discriminación en que esos sistemas pueden incurrir[88].

Se ha señalado que es este el primer esfuerzo legislativo federal para regular la IA en todas las industrias[89], propiciando una norma que procura que las entidades alcanzadas por la regulación realicen "evaluaciones de impacto" de sus sistemas de decisión automatizada cuando sea de "alto riesgo", de modo tal de evaluar los impactos que el sistema de IA puede tener desde la perspectiva de la seguridad, la equidad, el sesgo y la discriminación[90].

En el proyecto, un sistema se considera de alto riesgo si cumple alguna de estas condiciones[91]:

1. Plantee un riesgo significativo para la privacidad o seguridad de la información personal;
2. Pueda contribuir en decisiones injustas, inexactas, sesgadas o discriminatorias que afecten a consumidores;
3. Adopte decisiones o facilite la toma de decisiones, basado en una evaluación extensiva de y sistemática de los consumidores, incluyendo intentos de analizar o predecir aspectos sensibles de su vida -v. gr. su desempeño laboral, estado de salud, movimientos, situación económica-, que alteren derechos legales de consumidores o impacten significativamente en ellos;
4. Involucre información personal de un número significativo de consumidores como la raza, el color, la sexualidad, etc.;
5. Monitoree lugares de acceso público y grandes dimensiones;
6. Cumpla con cualquier otra condición que se incluya por vía reglamentaria.

El proyecto de ley determina que la Comisión Federal de Comercio (FTC, por sus siglas en inglés), será quien deberá resolver por vía reglamentaria los detalles necesarios para cumplir los objetivos de la ley[92].

Además, de acuerdo con el proyecto, la evaluación de impacto del sistema, cuando sea razonable, deberá ser

realizada con la intervención de terceros externos, independientes, expertos en tecnología o auditores⁹⁴; de forma previa cuando se trate de sistemas nuevos y con tanta frecuencia como la CFT determine cuando se trate de sistemas ya existentes^[93].

Las entidades alcanzadas por la regulación serán aquellas empresas que^[95]:

1. Facturen USD 50 millones o más al año;
2. Posean datos de más de 1 millón de consumidores o de los dispositivos de más de 1 millón de consumidores;
3. Actúen como intermediarios de datos que compran y venden información personal, independientemente de su facturación o del número de individuos cuyos datos posean.

La definición de las empresas que estarán alcanzadas por la norma en caso de que se apruebe, ha merecido la crítica de algunos sectores que la entienden restringida, en tanto las pequeñas empresas no alcanzadas pueden igualmente valerse de sistemas de IA que pueden provocar daños a las personas pero quedarán excluidas^[96].

Por otro lado, la propuesta determina que la evaluación de impacto del sistema de alto riesgo deberá analizar tanto al sistema como al proceso de desarrollo del sistema de decisión, incluyendo el diseño y el conjunto de datos de entrenamiento, en cuanto a sus impactos en materia de precisión, equidad, sesgo discriminación, privacidad y seguridad.

A tal fin, y cómo mínimo, deberá incluir^[97]:

1. Una descripción detallada del sistema, su diseño, datos de entrenamiento y propósito.
2. Una evaluación de los beneficios y costos del sistema a la luz de su propósito, considerando factores relevantes como el tiempo de guardado de información personal y resultados, qué información estará disponible para los consumidores, qué acceso tienen los consumidores a los resultados del sistema y cuáles son sus posibilidades de impugnar o corregir la decisión adoptada con él.
3. Una evaluación de los riesgos que plantea el sistema a la privacidad o seguridad de la información personal de los consumidores, y el riesgo de que las decisiones con él adoptadas puedan resultar inexactas, injustas, tendenciosas o discriminatorias.
4. Las medidas que la entidad alcanzada adoptará para minimizar esos riesgos, tanto tecnológicas como físicas.

El proyecto exige, asimismo, que las compañías concreten una evaluación de impacto de protección de datos para sus sistemas de información de alto riesgo, que evalúe hasta dónde el sistema protege la seguridad y privacidad de la información personal que procesa.

Se ha señalado que si bien el proyecto requiere a las compañías alcanzadas que evalúen la conformidad de sus algoritmos con ciertos estándares como la privacidad, seguridad y equidad, lo cierto es que no les exige revelar el código fuente o la fórmula del algoritmo de machine learning empleado, ni tampoco explicar cómo funcionan^[98].

Del mismo modo, tampoco exige que se hagan públicas las evaluaciones de impacto, siendo que ello permitiría al consumidor tener mayor información para ser más consciente de los posibles riesgos que conlleva interactuar con un sistema algorítmico determinado, y a su vez redundaría en una competencia entre las empresas para reducir ese riesgo^[99].

Finalmente, se ha criticado que la evaluación de impacto, no obstante ser una herramienta válida, no es suficiente por sí sola para lograr la responsabilidad algorítmica, con lo cual el marco regulatorio para la rendición de cuentas debería ser más amplio, incorporando una mayor variedad de mecanismos tecnológicos y de procedimiento para garantizar que el operador del sistema pueda corroborar que actúa de acuerdo a sus

objetivos, e identificar y corregir resultados perjudiciales[100].

V. Impugnar la decisión automatizada que discrimina - Las categorías sospechosas como herramienta

Hasta tanto existan normas que regulen la cuestión o surjan otras herramientas específicas de control, dado que la IA ya es presente y no sólo futuro, debemos bucear en el conjunto de preceptos y teorías que tenemos disponibles para dar con alguna herramienta que nos permita encauzar la impugnación de la decisión automatizada que provoca un tratamiento discriminatorio.

El primer aspecto a tener en cuenta es que en cualquier caso, aun cuando no haya un marco que regule su empleo, las aplicaciones de IA deben respetar los derechos fundamentales consagrados en los tratados de Derechos Humanos suscriptos por los distintos países, así como en sus constituciones, ya que esos marcos jurídicos delimitan el obrar de todas las personas y organizaciones, sea que adopten decisiones con o sin intervención de sistemas de IA.

En la actualidad, ese marco jurídico nos provee de una teoría que, entiendo, puede ser de suma utilidad para plantear y decidir un caso cuando se cuestiona una decisión automatizada porque presuntamente discrimina.

Me refiero aquí a la doctrina de las categorías sospechosas, de cuna estadounidense pero con un fructífero desarrollo en el ámbito europeo y latinoamericano (LATAM), que ha sido adoptada hace pocos años, no sólo por distintos países de nuestra región, sino por la propia Corte Interamericana de Derechos Humanos (CIDH).

V.1. Las categorías sospechosas en el ámbito latinoamericano y en particular, en el ámbito de la Corte Interamericana de Derechos Humanos

La doctrina de las categorías sospechosas, que fue definitivamente adoptada en Argentina con el precedente Hooft[101] de la Corte Suprema de Justicia de la Nación en el año 2005, y que se utiliza en otros países de LATAM como México, Colombia, Chile y Perú, también ha sido receptada por la CIDH que la ha empleado para decidir distintos casos vinculados con el derecho a la igualdad y a la no discriminación[102].

En el ámbito latinoamericano, recordemos, el art. 1 de la CADH prohíbe discriminar por motivos de raza, color, sexo, idioma, religión, opiniones políticas o de cualquier otra índole, origen nacional o social, posición económica, nacimiento o cualquier otra condición social[103].

A la luz de este precepto, la CIDH ha señalado que existe un vínculo indisoluble entre la obligación de respetar y garantizar los derechos humanos, y el principio de igualdad y no discriminación, lo que determina el deber de los Estados de respetar y garantizar sin discriminación los derechos contenidos en la CADH[104].

Bajo esta premisa, ha receptado la doctrina de las categorías sospechosas en los precedentes Atala Riffo y niñas vs. Chile[105] y Norín Catrimán y otros vs. Chile[106], de los años 2012 y 2014, respectivamente, concluyendo que cuando estamos frente una distinción realizada por alguno de los motivos vedados por el art. 1 de la Convención Americana sobre Derechos Humanos, estaremos frente a lo que se conoce como una categoría sospechosa de discriminación.

Como consecuencia de ello, en estos casos se disparará un control judicial de excepción, que exigirá al Estado una fundamentación rigurosa, trasladándose a su cargo el peso de probar que la decisión no tenía un propósito, ni un efecto discriminatorio[107].

En otras palabras, la presencia de un motivo prohibido o categoría sospechosa exigirá una fundamentación rigurosa y de mucho peso a quien pretenda sostener la diferencia de trato, y en virtud de la inversión de la carga probatoria que trae su presencia, le demandará al acusado demostrar que su decisión no tenía un propósito ni un efecto discriminatorio[108].

Desde estos precedentes, la CIDH ha empleado esta doctrina para juzgar diferencias de trato en razón del origen étnico[109], la orientación sexual[110], la opinión política[111], la pobreza como condición económica[112] y la

condición de portador de HIV^[113], sentando que el carácter sospechoso de la norma que las utilice vendrá dado por el hecho de que la diferencia de trato se basa en alguno de los motivos de discriminación prohibidos por el art. 1.1. de la Convención Americana sobre Derechos Humanos que, recuerdo, vedan discriminar con motivo en la raza, el color, el sexo, el idioma, la religión, las opiniones políticas o de cualquier otra índole, el origen nacional o social, la posición económica, el nacimiento o cualquier otra condición social.

V.2. Categorías sospechosas y potencialmente sospechosas - Relevancia de la distinción

La doctrina a la que se viene haciendo referencia tiene origen en la Nota al Pie 4 del precedente *Carolene Products* de la Suprema Corte de los Estados Unidos, en el que se justificó por primera vez la posibilidad de que el Poder Judicial controle más severamente ciertas elecciones hechas por el legislador al regular, dejando de lado el control de mera razonabilidad que aplica como regla, con el fin de dar una protección especial a determinados grupos que se consideran vulnerables, cuando se ven afectados por las normas^[114].

¿Cuál es la característica distintiva de estos grupos que los hace merecedores de una tutela judicial diferenciada? Que son minoritarios desde una perspectiva política, a partir de que denotan dificultades para participar efectivamente de un proceso político que por regla es mayoritario, y que por eso no los protege del mismo modo en que resguarda a los grupos dominantes que son los que, en definitiva, adoptan esas reglas^[115].

Esta circunstancia es la que justifica que los jueces sean llamados a intervenir en el proceso político por medio de un control de constitucionalidad de excepción (conocido en los EE.UU como test estricto), dejando de lado el test de razonabilidad que normalmente aplican para examinar la constitucionalidad de las leyes, sin que esta opción vulnere la división de poderes^[116].

Considerando lo expuesto, he concluido en otra oportunidad que para la adecuada identificación de las categorías sospechosas se necesita, primeramente, corroborar que la diferencia de trato alegada como discriminatoria se basa en alguno de los motivos de discriminación prohibidos por la CADH, en cuyo caso estaremos en presencia de una categoría "potencialmente sospechosa"^[117].

Luego, la naturaleza "sospechosa" la distinción se confirmará si la diferencia de trato afecta a un grupo que carece de herramientas para solucionar la situación de desventaja en la que se encuentra mediante una participación efectiva en el proceso político que, por regla, es mayoritario y que no la protege del mismo modo en que resguarda a los grupos predominantes, que sí están bien representados en dicho sistema^[118].

Para identificar estos grupos existen instrumentos específicos, que se han bautizado bajo el nombre de "indicios" o "factores" de sospecha, que van a estar presentes con relación al grupo perjudicado por la diferencia de trato, y que serán los que van a permitir confirmar el carácter sospechoso de la distinción.

Estos indicios vienen siendo empleados, tanto por el Tribunal Europeo de Derechos Humanos, como por diferentes tribunales superiores de LATAM, como el Tribunal Constitucional de Colombia, que emplean esta tesis para decidir casos en involucran el derecho a la igualdad y a la no discriminación^[119].

Constituyen indicios de sospecha, a modo de ejemplo, que la persona integre una clase que históricamente ha sufrido trato discriminatorio (v.gr. mujeres, personas con discapacidad, colectivo LGTB); que la distinción recaiga sobre una característica de esa persona que es innata e inmodificable (v.gr. la raza); si se comprueba que el distingo refleja estereotipos negativos del grupo que están arraigados socialmente pero deben modificarse (v.gr. como ha sucedido con regulaciones que afectan a los adultos mayores o a los niños, y están vinculadas con su presunta falta de autonomía para tomar decisiones)^[120].

En suma, el carácter sospechoso de una diferencia de trato se confirmará siempre que estemos frente a una distinción realizada sobre la base de un motivo vedado que, además, afecte negativamente a un grupo con relación al cual identificamos la presencia de uno o varios indicios de sospecha^[121].

Y en estos casos, se disparará la alternativa de que distinción que los emplea se examine bajo un test severo, que va a determinar su presunción de inconstitucionalidad e invertirá la carga de la prueba, exigiendo a quien

pretenda sostener su validez una fundamentación rigurosa y de mucho peso, así como demostrar que su decisión no tenía un propósito ni un efecto discriminatorio[122].

V.3. Las categorías sospechosas para impugnar y controlar el sesgo algorítmico

Si un sistema de IA es entrenado con datos que presenten algún tipo de sesgo (racial, de género, por la condición socio económica, etc.), los resultados generados por ese sistema no sólo van a reproducir el mismo sesgo, sino que es probable que lo amplíen y lo institucionalicen en su ámbito de aplicación[123].

El potencial de la IA de establecer y perpetuar sesgos y discriminación[124], demanda un gran trabajo interdisciplinario, en el que operadores informáticos y jurídicos deberán colaborar para desarrollar herramientas que permitan identificar los sesgos para prevenir sus resultados indeseados[125].

Desde el derecho se ha postulado la vigencia de una serie de principios jurídicos generales dirigidos a lograr una IA que sea compatible con el Estado constitucional de Derecho y con los Derechos Humanos. Entre ellos se destaca, sin dudas, el principio de no discriminación algorítmica[126].

En virtud de este principio general, debe exigirse que el diseño y la implementación de sistemas de IA respeten el principio de no discriminación, lo que impide aceptar que estos sistemas procesen la información o los datos de las personas humanas sobre la base de criterios de distinción prohibidos, como puede suceder con la raza, la posición económica y la religión, entre otros[127].

En línea con ello, si el resultado de la aplicación de un sistema de IA se explica por categorías prohibidas nos encontraremos frente a un supuesto de discriminación algorítmica y el sujeto afectado podrá acudir a la justicia para solicitar que se deje sin efecto el acto o decisión discriminatorio[128].

Ahora bien, la alternativa de demandar eficazmente el cese del acto discriminatorio se encuentra directamente relacionada con el principio de explicabilidad algorítmica ya referido más arriba[129], cuya efectiva vigencia será sustancial para que el sujeto afectado por la decisión asociada a un sistema de IA pueda ejercer su derecho de defensa en oportunidad de cuestionarla[130].

Considerando este extremo, debemos tener en cuenta que los diversos enfoques de aprendizaje automático presentan (al menos al momento de escribir estas líneas) enormes diferencias en orden a su explicabilidad, y estas diferencias deben ponderarse al tiempo de diseñar la estrategia defensiva dirigida a impugnar una decisión automatizada por presumirla discriminatoria[131].

Es que el resultado de un árbol de decisiones, que es un típico ejemplo de IA de caja blanca, posiblemente se pueda explicar enteramente a través de la secuencia de *test* que llevan a cierto resultado[132], lo que no sólo va a facilitar la tarea de formular la impugnación al afectado, sino la del juez de corroborar si la decisión asociada al sistema de IA es o no discriminatoria.

En cambio, cuando se trata de resultados provenientes de la aplicación de sistemas de caja negra, por la opacidad propia de la técnica, la tarea de identificar y probar que un determinado resultado obedece a un distingo prohibido se volverá ciertamente más ardua y en ocasiones, imposible[133].

No sólo porque estos sistemas suelen estar protegidos por regulaciones de propiedad intelectual que impiden el acceso al código (aunque esto puede darse también con la IA de caja blanca), sino además porque por la propia dinámica de funcionamiento de esta clase de redes, no es posible explicar totalmente el paso a paso que permita comprender cómo se han sopesado los diversos atributos y la relevancia que se ha asignado a cada dato e información para arribar a un resultado[134].

Sumado a ello, existen algoritmos que pueden optimizar su funcionamiento de forma continuada, es decir, que aprenden y se refinan con el tiempo, lo que lleva a que los parámetros que emplea el algoritmo y la forma de arribar a los resultados varíe de un momento a otro, tornando aún más complejo resolver la problemática vinculada a su explicabilidad[135].

Dada esta realidad, considero que cuando una persona se ve perjudicada por una decisión y sospeche que ella puede ser el resultado de, o estar basada en, la aplicación de un sistema de IA que ha utilizado motivos de distinción prohibidos, sea directa o indirectamente, dando lugar a un resultado presuntamente discriminatorio o injusto, se justificaría apelar a la doctrina de las categorías sospechosas, que puede convertirse en una gran aliada para controlar y juzgar el sesgo algorítmico.

Apelar a esta solución, que trasladará al acusado la carga de probar que el trato no es discriminatorio, se justifica por dos motivos:

1. Por un lado, en que las dificultades del afectado para probar el trato discriminatorio van a estar directamente relacionadas con la opacidad propia de los sistemas de IA de caja negra o bien, con las razones de propiedad intelectual esgrimidas para negar el acceso al código por quien los emplea, lo que lo coloca en una situación de inferioridad con relación a que se vale del sistema, tornando irrazonable que sea a él a quien se le exija probar cómo se arribó a la decisión.

2. Considerando lo anterior, sería razonable que quien ha elegido emplear el sistema de IA, que es quien se beneficia en mayor medida con su utilización, sea quien asuma la carga de probar que el resultado o la decisión impugnada no obedece a un trato discriminatorio.

No se me pasa por alto que esta doctrina fue inicialmente pensada para examinar diferencias de trato contenidas en normas, pero aun así, dado el loable objetivo que persigue, consistente en proveer de una tutela judicial diferenciada a grupos vulnerables, y atendiendo a su indudable utilidad, determinada por la presunción de invalidez que trae la sola presencia de una categoría sospechosa, la consecuente inversión de la carga probatoria y la activación de un control judicial severo; su empleo podría hacerse extensivo a los casos de discriminación algorítmica cuando ella provenga del empleo de un sistema de IA por parte de un particular o del propio Estado (aunque este último no debería poder basar sus decisiones en sistemas de IA de caja negra[136]).

Todo ello en tanto los Estados tienen el deber de garantizar a las personas sujetas a su jurisdicción, el libre y pleno ejercicio de los derechos y libertades reconocidos en la CADH, sin discriminación alguna por motivos de raza, color, sexo, idioma, religión, opiniones políticas o de cualquier otra índole, origen nacional o social, posición económica, nacimiento o cualquier otra condición social[137]; y al menos hasta tanto se desarrollen instrumentos jurídicos o informáticos más específicos para atender esta problemática.

De aceptarse esta propuesta, para el sujeto afectado negativamente por el resultado de la aplicación de un sistema de IA, debería bastar con alegar que existen indicios serios de que el trato discriminatorio puede basarse en una categoría vedada y que pertenece a una clase vulnerable en los términos señalados más arriba, solicitando el análisis de su caso bajo los parámetros de esta doctrina[138].

Como consecuencia de ello, se trasladará a quien ha elegido valerse del sistema de IA, la carga de probar que la decisión o el resultado se basa en distinciones lícitas o bien, que la diferencia de trato se justifica en razones de mucho peso y que no tenía un propósito ni un efecto discriminatorio[139].

Ello independientemente de que quien utilice el sistema de IA lo emplee como soporte o que le hubiera delegado directamente la decisión, pues insisto, es quien en definitiva lo ha elegido y el que más aprovecha de los beneficios que estos sistemas suelen reportar (mayor eficiencia, agilidad, eficacia y menores costos, entre otros)[140].

VI. Conclusiones

Los sistemas de IA deben ser compatibles con los principios del Estado Constitucional de Derecho y con los Derechos Humanos[141]. Una IA con un enfoque centrado en la persona deberá esforzarse por asegurar que los valores humanos tengan un papel central "...en el desarrollo, despliegue, utilización y supervisión de los sistemas de IA, garantizando el respeto de los derechos fundamentales..."[142].

La IA debe centrarse en las personas y fundarse en el compromiso de ser utilizada al servicio de la humanidad y del bien común, con el objetivo de mejorar el bienestar común y traer consigo progreso e innovación[143].

En ese marco, la IA debe ser lícita, debe de ser ética y debe ser robusta, tanto desde el punto de vista técnico como social[144].

No debe perderse de vista, además, que "...los seres humanos desempeñan un papel fundamental en el diseño y la difusión de sistemas de IA, definiendo los objetivos de una aplicación de IA y, según el tipo de aplicación, eligiendo y etiquetando conjuntos de datos y clasificando productos (...) siempre determinan la aplicación y el uso de los productos de la IA, incluido el grado en que complementan o reemplazan la adopción de decisiones humana..."[145].

La dimensión sustantiva del principio de equidad en materia de sistemas de IA[146], así como el principio de no discriminación algorítmica[147], aseguran que ninguna persona sufra sesgos injustos, discriminación, ni estigmatización, constituyéndose en un claro límite frente a algoritmos que parecen ser neutros cuando en realidad no lo son[148].

Aun así, debe ponerse especial atención a aquellas situaciones que afecten grupos vulnerables, como son los niños, las mujeres, las personas con discapacidad, los adultos mayores y otros colectivos que históricamente se han visto desfavorecidos o se encuentran en riesgo de exclusión, pues los sesgos presentes en los datos con los que se entrena el sistema de IA pueden terminar por colocarlos en una espiral descendente de mayor desigualdad y exclusión[149].

Es que como hemos señalado, los sistemas de IA pueden estar entrenados con prejuicios del pasado y ello llevará a que su aplicación no haga más que reforzarlos con serias chances de que el sistema, al basarse en esos datos del pasado para adoptar decisiones en el presente, reproduzcan la lógica (sexista, racista o prejuiciosa) insita en los datos, colaborando en el mantenimiento o provocando una mayor exclusión de grupos tradicionalmente desventajados.

Además, debemos tener presente que en ocasiones es muy arduo desafiar el trato injusto provocado por decisiones automatizadas, ya que los cuestionamientos realizados por personas interesadas que son conscientes de esta problemática, aun cuando estén justificados, suelen ser ignorados o rechazados con fundamento en que interfieren en el funcionamiento del sistema[150].

A ello se agrega que frente a decisiones adoptadas con sistemas de IA de caja negra, para el individuo afectado por la decisión será muy engorroso impugnarla y defenderse adecuadamente de ella, dada la problemática de su poca explicabilidad referida más arriba. Lo mismo sucede en los casos en que se argumentan motivos de propiedad industrial o intelectual para negar el acceso al código fuente[151].

De hecho, la naturaleza opaca de los algoritmos de caja negra ya enciende alarmas respecto de su empleo en ciertas áreas sensibles, como aquellas que hacen a nuestra vida política, social o económica, en las que posiblemente queramos plantearnos como sociedad que no exista lugar para las decisiones basadas en algoritmos[152].

Conscientes de los beneficios que si dudas traerá la IA, pero alertados también sobre los riesgos que conlleva para las personas su desarrollo, implementación y empleo, los Estados y distintas organizaciones *supra* nacionales, han dado inicio a un saludable debate en pos de diseñar los marcos regulatorios que permitan aprovechar y promover la utilización de estos sistemas, en paralelo que garanticen que sean seguros, que contribuyan al bienestar y al interés general de la población y que sean compatibles con el Estado Constitucional de Derecho y respetuosos de los derechos fundamentales reconocidos a las personas humanas.

En ese marco, existen distintas propuestas para regular los sistemas de IA artificial, que propician distintos enfoques, algunos de los cuales se basan mayormente en la consagración de diversos principios para el logro de una IA ética, robusta y fiable, y otros en el establecimiento de reglas concretas, así como de propuestas dirigidas a perseguir la rendición de cuentas algorítmica por parte de las empresas que se valen de sistemas que utilizan

IA y cuya aplicación es considerada de alto.

No obstante, en la mayoría de los países de LATAM este es un debate aún nos debemos como sociedad, que suele verse postergado por otras urgencias que -entendiblemente- ocupan la agenda política y social, como la pobreza, la macroeconomía y la educación, más que debe incorporarse necesariamente si queremos dar adecuada tutela a los usuarios y consumidores, y a la población en general, frente al avance en el desarrollo e implementación de sistemas de IA en nuestras latitudes.

Sin perjuicio de ello, lo cierto es que hasta tanto tengamos un marco regulatorio que nos permita cuestionar e impugnar la decisión automatizada que ha sido adoptada con apoyo en sistemas de IA, o bien por el propio sistema, no nos queda otra opción que apelar a las herramientas actualmente disponibles en nuestro sistema jurídico, pues cuando esas decisiones son sesgadas, la persona afectada no puede quedar sin respuesta.

Aquí, entiendo, la doctrina de las categorías sospechosas, receptada en el ámbito latinoamericano tanto a nivel interno de varios países, como a nivel del Sistema Interamericano de Derechos Humanos, aparece como un instrumento que puede ser muy apropiado y útil para impugnar la decisión discriminatoria y controlar el sesgo algorítmico.

Esta teoría habilitará que frente a un supuesto de discriminación algorítmica, siempre que el afectado pueda dar cuenta de indicios serios que hagan presumir que es víctima de un trato discriminatorio, el resultado de la aplicación del sistema se juzgue a su amparo, lo que determinará que se traslade a quien se beneficia del empleo del sistema la carga de probar que la decisión o el resultado se basa en distinciones lícitas o bien, que la diferencia de trato se justifica en razones de mucho peso y que no tenía un propósito ni un efecto discriminatorio^[153].

Considero que esta alternativa se presenta no sólo como una solución posible, sino también razonable, en aras de hacer efectivo el derecho de defensa del afectado, así como su derecho a la igualdad, reflejado en el principio de no discriminación algorítmica. Más aún cuando el afectado haya sido objeto de una decisión automatizada adoptada con, o en base a, sistemas de IA de caja negra^[154].

VII. Bibliografía

Algorithmic Accountability Act of 2019. <https://www.congress.gov/bill/116th-congress/house-bill/2231/all-info>.

Allen Robin QC y Masters Dee. "*Algorithms, Apps & Artificial Intelligence: The next frontier in discrimination law*". Publica Law Project Annual Conference, Octubre-2018. <https://www.cloisters.com/wp-content/uploads/2018/10/PLP-Apps-Algorithms-and-AI-paper-final.pdf>.

Allen Robin QC y Masters Dee, "*In the matter of automated data processing in government decision making*", September-2019, open opinion for The Legal Education Foundation, <https://www.cloisters.com/wp-content/uploads/2019/10/Open-opinion-pdf-version-1.pdf>.

Amunátegui Perelló, Carlos y Madrid Raúl. "*Sesgo e Inferencia en Redes Neuronales ante el Derecho*", en C. Aguerre (Ed.), *Inteligencia Artificial en América Latina y el Caribe. Ética, Gobernanza y Políticas*, Bs. As., CETyS Universidad de San Andrés, 1-22, 2020. <https://guia.ai/documentos-2020/>.

Asamblea General ONU. Resolución 73/348. <http://undocs.org/es/A/73/348>.

Barredo, Alex, "*Google, Amazon, Facebook y Microsoft se unen para que la inteligencia artificial salga a la calle*", La Vanguardia, 1/10/2016, <https://www.lavanguardia.com/tecnologia/20161001/41684797831/inteligencia-artificial-google-facebook-amazon-apple.html>.

Chae Yoon, "U.S. AI Regulation Guide: Legislative Overview and Practical Considerations", en *RAIL The Journals of Robotics, Artificial Intelligence and Law* Vol. 3 N° 1, Jan-Feb, 17-40, 2020, Full Court Press, Washington.

Comisión Europea. High-Level Expert Group on Artificial Intelligence, *A definition of AI: Main capabilities and scientific disciplines*, 2019. <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>.

Comisión Europea. *Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza*. Bruselas, 2020.

https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf.

Comisión Europea, *Directrices para una IA fiable, Grupo de Expertos de alto nivel sobre inteligencia artificial*, Bruselas, 2018. <https://op.europa.eu/en/publication-detail/->

</publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es/format-PDF>.

Corte IDH. "Trabajadores de la Hacienda Brasil Verde vs. Brasil", 20/10/2016. https://www.corteidh.or.cr/docs/casos/articulos/seriec_318_esp.pdf.

Corte IDH. "González Lluy y otros vs. Ecuador", 1/9/2015. https://www.corteidh.or.cr/docs/casos/articulos/seriec_298_esp.pdf.

Corvalán, Juan G. "Inteligencia Artificial, Automatización y Predicciones en el Derecho". Artículo en prensa que formará parte de la obra, Dupuy Daniela y Corvalán Juan G. (Dir.), *Cibercrimen III, Inteligencia Artificial, Automatización, algoritmos y predicciones y en el Derecho Penal*, IbdeF Montevideo-Buenos Aires, 2020.

Corvalán, Juan G. *Perfiles Digitales Humanos*. Buenos Aires, L.L., 2020.

Dastin Jeffrey. "Amazon scraps secret AI recruiting tool that showed bias against women". Reuters, 10 de octubre de 2018. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-thatshowed-bias-against-women-idUSKCN1MK08G>.

European Union Agency for Fundamental Rights. *Getting de future right. Artificial Intelligence and fundamental rights*. Luxemburgo, 2020. <https://fra.europa.eu/en/publication/2020/artificial-intelligence-and-fundamental-rights>.

"Facebook, Amazon, Google, IBM y Microsoft se unen para crear una organización de IA", 30/9/2016, <https://medium.com/@W3Americas/facebook-amazon-google-ibm-y-microsoft-se-unen-para-crear-la-asociaci%C3%B3n-de-ia-89efd92f8ecc>

Centre for Data Ethics and Innovation. *Review into bias in algorithmic decision-making*. U.K., 2020. <https://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making>.

Gascón Marcer, Ana. "Derechos Humanos e Inteligencia Artificial", en Pérez Mirás A., Tarufel Lozano Germán M., Raffiotta Eduardo, Iadidicco María Pía (Dir.), *Setenta Años de Constitución Italiana y Cuarenta Años de Constitución Española*, Vol. V. Centro de Estudios Políticos y Constitucionales, 335-350, 2020.

https://www.researchgate.net/publication/339687454_Derechos_Humanos_e_Inteligencia_Artificial.

"Google pide perdón a una pareja por confundirla con gorilas". *BBC Mundo Tecnología*, 2 de julio de 2015.

https://www.bbc.com/mundo/noticias/2015/07/150702_tecnologia_google_perdon_confundir_afr_oamericanos_gorilas_lv.

Hao, Karen. "This is how AI bias really happens – and why it's so hard to fix". *MIT Technology Review*, 4 de febrero de 2019. https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/?utm_medium=tr_social&utm_campaign=site_visitor.unpaid.engagement&utm_source=Link

edIn#Echobox=1610662231.

lenca Marcelo y Vayena Effy. "AI Ethics Guidelines: European and Global Perspectives", en CAHAI Secretariat (Comp.), *Towards Regulation of AI Systems*, 38-60, 2020. <https://www.coe.int/en/web/artificial-intelligence/-/-toward-regulation-of-ai-systems->.

Martín García Sergio. "Ética e inteligencia artificial". *Cuadernos de la Cátedra CaixaBank de Responsabilidad Social Corporativa*, no. 42 (septiembre 2019): 4-25. <https://www.iese.edu/wp-content/uploads/2019/12/ST-0522.pdf>.

MacCarthy Marck. *An examination of the Algorithmic Accountability Act*, 2019. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3615731.

Moisés Barrio, Andrés, "Por qué es necesario regular la inteligencia artificial", *Confilegal*, (2019), <https://confilegal.com/20190117-por-que-es-necesario-regular-la-inteligencia-artificial/>.

Muller Catelijne. "The Impact of IA on Human Rights, Democracy and the Rule of Law", en CAHAI Secretariat (Comp.), *Towards Regulation of AI Systems*, 21-37, 2020. <https://www.coe.int/en/web/artificial-intelligence/-/-toward-regulation-of-ai-systems->.

Parlamento Europeo, *El impacto del Reglamento General de Protección de Datos (GDPR) en la inteligencia artificial*, Bruselas, 2020.

[https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU\(2020\)641530_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf).

Parlamento Europeo. *Marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas*, 2020. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_ES.pdf.

Sánchez Caparrós, *Categorías Sospechosas*, Astrea, Bs. As., 2020

Sexismo en la inteligencia artificial. <https://www.youtube.com/watch?v=PkBdBDwVvJs>.

Zender, Daniel. "The coming war on the hidden algorithms that trap people in poverty". MIT Technology Review, 4 de diciembre de 2020.

<https://www.technologyreview.com/2020/12/04/1013068/algorithms-create-a-poverty-trap-lawyers-fight-back/>.

[1]

Centre for Data Ethics and Innovation, Review into bias in algorithmic decision-making (U.K., 2020), p. 3, <https://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making> (Consultado el 29/06/2021).

[2]

Comisión Europea, Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza (Bruselas, 2020), p. 14, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf.

[3]

Zender, Daniel, "The coming war on the hidden algorithms that trap people in poverty", MIT Technology Review, (Diciembre-2020), <https://www.technologyreview.com/2020/12/04/1013068/algorithms-create-a-poverty-trap-lawyers-fight-back/> (Consultado el 29/06/2021).

[4]

Ídem.

[5]

Allen Robin QC y Masters Dee, "Algorithms, Apps & Artificial Intelligence: The next frontier in discrimination law", Publica Law Project Annual Conference, (Octubre-2018), p. 12, <https://www.cloisters.com/wp-content/uploads/2018/10/PLP-Apps-Algorithms-and-AI-paper-final.pdf>.

[6]

Ibidem, p. 12.

[7]

"Sexismo en la inteligencia artificial", <https://www.youtube.com/watch?v=PkBdBDwVvJs> (Consultado el 29/06/2021).

[8]

Ver Dustin Jeffrey. "Amazon scraps secret AI recruiting tool that showed bias against women". Reuters, (10/10/2018), <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-thatshowed-bias-against-women-idUSKCN1MK08G>. También "Google pide perdón a una pareja por confundirla con gorilas", BBC Mundo Tecnología (2/7/2015) (Consultado el 29/06/2021), https://www.bbc.com/mundo/noticias/2015/07/150702_tecnologia_google_perdon_confundir_afroamericanos_gorilas_lv; y Hao, Karen. "This is how AI bias really happens - and why it's so hard to fix". MIT Technology Review, (4/2/2019), https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/?utm_medium=tr_social&utm_campaign=site_visitor.unpaid.engagement&utm_source=LinkedIn#Echobox=1610662231 (Consultado el 29/06/2021).

[9]

Asamblea General ONU, Resolución 73/348, p. 5, <http://undocs.org/es/A/73/348>.

[10]

Ídem.

[11]

Martín García Sergio, "Ética e inteligencia artificial", Cuadernos de la Cátedra CaixaBank de Responsabilidad Social Corporativa, N° 42 (Septiembre-2019), p. 7, <https://www.iese.edu/wp-content/uploads/2019/12/ST-0522.pdf> (Consultado el 29/06/2021). También European Union Agency for Fundamental Rights, Getting de future right. Artificial Intelligence and fundamental rights, (Luxemburgo, 2020), p. 19, <https://fra.europa.eu/en/publication/2020/artificial-intelligence-and-fundamental-rights>.

[12]

Comisión Europea, Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza, (Bruselas, 2020), p. 21, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf.

[13]

Asamblea General ONU, Resolución 73/348, p. 4, <http://undocs.org/es/A/73/348>. Ver también: Parlamento Europeo, El impacto del Reglamento General de Protección de Datos (GDPR) en la inteligencia artificial, (2020), p. 3, [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU\(2020\)641530_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf) (Consultado el 29/06/2021).

[14]

Parlamento Europeo, El impacto..., p. 3, [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU\(2020\)641530_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf) (Consultado el 29/06/2021).

[15]

Corvalán, Juan G., Perfiles Digitales Humanos, Buenos Aires, L.L., 2020, p. 33.

[16]

Asamblea General ONU, Resolución 73/348, p. 4, <http://undocs.org/es/A/73/348>. También Comisión Europea, High-Level Expert Group on Artificial Intelligence, A definition of AI: Main capabilities and scientific disciplines, (2019), p. 5, <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines> (Consultados el 29/06/2021).

[17]

Corvalán, Juan G., Perfiles..., p. 37/39.

[18]

Corvalán, Juan G., Perfiles..., p. 37/39. También Comisión Europea, High-Level Expert Group on Artificial Intelligence, A definition of AI..., p. 5, <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines> (Consultado el 29/06/2021).

[19]

Corvalán, Juan G., "Inteligencia Artificial, Automatización y Predicciones en el Derecho". Artículo en prensa que formará parte de la obra, Dupuy Daniela y Corvalán Juan G. (Dir.), Cibercrimen III, Inteligencia Artificial, Automatización, algoritmos y predicciones y en el Derecho penal, IbdeF Montevideo-Buenos Aires, 2020, p. 11.

[20]

Comisión Europea, High-Level Expert Group on Artificial Intelligence, A definition of AI..., p. 5, <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines> (Consultado el 29/06/2021).

[21]

lenca Marcelo y Vayena Effy, "AI Ethics Guidelines: European and Global Perspectives", en CAHAI Secretariat (Comp.), Towards Regulation of AI Systems, (2020), p. 51, <https://www.coe.int/en/web/artificial-intelligence/-/toward-regulation-of-ai-systems->.

[22]

Comisión Europea, Grupo de Expertos de alto nivel sobre inteligencia artificial, Directrices para una IA fiable, (2018), p. 16, <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es/format-PDF>.

[23]

Ibídem, p. 29.

[24]

Corvalán, Juan G., *Perfiles...*, p. 136. También, Corvalán, Juan G., "Inteligencia Artificial, Automatización y Predicciones en el Derecho"..., p. 48.

[25]

Red Europea de Expertos Legales en igualdad de género y no discriminación, *Discriminación algorítmica en Europa*, Bruselas, 2020, p. 49.

[26]

Ídem.

[27]

Ídem.

[28]

Ibídem, p. 49/50.

[29]

Comisión Europea, Grupo de Expertos de alto nivel sobre inteligencia artificial, *Directrices...*, p. 51, <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es/format-PDF> (Consultado el 29/06/2021).

[30]

Asamblea General, ONU, Resolución 73/348, p. 5 disponible en <http://undocs.org/es/A/73/348>.

[31]

Comisión Europea, *Directrices...*, p. 25, <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es/format-PDF>.

[32]

Asamblea General, ONU, Resolución 73/348, p. 15, <http://undocs.org/es/A/73/348> (Consultado el 29/06/2021).

[33]

Amunátegui Perelló, Carlos y Madrid Raúl, "Sesgo e Inferencia en Redes Neuronales ante el Derecho", en C. Aguerre (Ed.), *Inteligencia Artificial en América Latina y el Caribe. Ética, Gobernanza y Políticas*, Bs. As., CETyS Universidad de San Andrés, (2020), p. 5, <https://guia.ai/documentos-2020/> (Consultado el 29/06/2021).

[34]

Ídem.

[35]

Parlamento Europeo, El impacto..., p. 21, https://www.europarl.europa.eu/stoa/en/document/EPRS_STU%282020%29641530 (Consultado el 29/06/2021).

[36]

Ídem.

[37]

Ibidem, p. 20.

[38]

Ídem.

[39]

Ídem.

[40]

Gascón Marcer, Ana, "Derechos Humanos e Inteligencia Artificial", en Pérez Mirás A., Tarufel Lozano Germán M., Raffiotta Eduardo, Iadidicco María Pía (Dir.), Setenta Años de Constitución Italiana y Cuarenta Años de Constitución Española, Vol. V, Centro de Estudios Políticos y Constitucionales, (2020), p. 337, https://www.researchgate.net/publication/339687454_Derechos_Humanos_e_Inteligencia_Artificial (Consultado el 29/06/2021).

[41]

Corvalán, Juan G., Perfiles..., p. 137.

[42]

Muller Catelijne, "The Impact of IA on Human Rights, Democracy and the Rule of Law", en CAHAI Secretariat (Comp.), ob. cit., (2020), p. 24, <https://www.coe.int/en/web/artificial-intelligence/-/toward-regulation-of-ai-systems-> (Consultado el 29/06/2021).

[43]

Allen Robin QC y Masters Dee, "Algorithms, Apps...", p. 8, <https://www.cloisters.com/wp-content/uploads/2018/10/PLP-Apps-Algorithms-and-AI-paper-final.pdf>.

[44]

Asamblea General, ONU, Resolución 73/348, p. 6, <http://undocs.org/es/A/73/348> (Consultado el 29/06/2021).

[45]

Corvalán, Juan G., Inteligencia Artificial y Derechos Humanos, Parte II, "Diario Constitucional y Derechos Humanos", Nro. 157, (2017), p. 2, <https://dpicuantico.com/sitio/wp-content/uploads/2017/07/Juan-Gustavo-Corvalan-Constitucional-10.07.2017.pdf>.

[46]

Martín García Sergio, ob. cit., p. 19, <https://www.iese.edu/wp-content/uploads/2019/12/ST-0522.pdf> (Consultado el 29/06/2021).

[47]

Ídem.

[48]

Comisión Europea, Directrices..., p. 14/15, <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es/format-PDF>. También Corvalán, Juan G., Inteligencia Artificial y Derechos Humanos..., p. 3, <https://dpicuantico.com/sitio/wp-content/uploads/2017/07/Juan-Gustavo-Corvalan-Constitucional-10.07.2017.pdf>.

[49]

Comisión Europea, Directrices..., p. 15, <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es/format-PDF>. También Corvalán, Juan G., Inteligencia Artificial y Derechos Humanos..., p. 8, <https://dpicuantico.com/sitio/wp-content/uploads/2017/07/Juan-Gustavo-Corvalan-Constitucional-10.07.2017.pdf>.

[50]

Comisión Europea, Directrices..., p. 15/16, <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es/format-PDF>. También Corvalán, Juan G., Inteligencia Artificial y Derechos Humanos..., p. 10, <https://dpicuantico.com/sitio/wp-content/uploads/2017/07/Juan-Gustavo-Corvalan-Constitucional-10.07.2017.pdf>.

[51]

Comisión Europea, Directrices..., p. 16, <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es/format-PDF>.

[52]

Comisión Europea, Directrices..., p. 23, <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es/format-PDF>.

[53]

Comisión Europea, Directrices..., p. 23, <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es/format-PDF>.

[54]

Comisión Europea, Directrices..., p. 23, <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es/format-PDF>.

[55]

Barredo, Alex, "Google, Amazon, Facebook y Microsoft se unen para que la inteligencia artificial salga a la calle", La Vanguardia, 1/10/2016, <https://www.lavanguardia.com/tecnologia/20161001/41684797831/inteligencia-artificial-google-facebook-amazon-apple.html>. También "Facebook, Amazon, Google, IBM y Microsoft se unen para crear una organización de IA", 30/9/2016, <https://medium.com/@W3Americas/facebook-amazon-google-ibm-y-microsoft-se-unen-para-crear-la-asociaci%C3%B3n-de-ia-89efd92f8ecc> (Consultado el 29/06/2021).

[56]

Moisés Barrio, Andrés, "Por qué es necesario regular la inteligencia artificial", Conflegal, (enero, 2019), <https://conflegal.com/20190117-por-que-es-necesario-regular-la-inteligencia-artificial/> (Consultado el 29/06/2021).

[57]

Ídem.

[58]

Comisión Europea, Libro Blanco sobre la inteligencia artificial, (Bruselas, 2020), p. 13/14, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf (Consultado el 29/06/2021).

[59]

Ibídem, p. 13 y 21.

[60]

Ibídem, p. 13.

[61]

Ibídem, p. 21.

[62]

Ibídem, p. 21.

[63]

Ibídem, p. 21.

[64]

Ibídem, p. 21.

[65]

Ibídem, p. 21.

[66]

Ibídem, p. 21.

[67]

Ibídem, p. 21.

[68]

Ibídem, p. 22.

[69]

Ibídem, p. 22/23.

[70]

Ibídem, p. 23/24.

[71]

Ibídem, p. 24.

[72]

Ibídem, p. 24.

[73]

Ibídem, p. 25.

[74]

Ibídem, p. 25/26.

[75]

Ibídem, p. 27.

[76]

"El Parlamento muestra el camino para la normativa sobre inteligencia artificial", Noticias del Parlamento Europeo, <https://www.europarl.europa.eu/news/es/press-room/20201016IPR89544/el-parlamento-muestra-el-camino-para-la-normativa-sobre-inteligencia-artificial> (Consultado el 29/06/2021).

[77]

Ídem.

[78]

Parlamento Europeo, Marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas, (octubre, 2020), p. 8, https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_ES.pdf (Consultado el 29/06/2021).

[79]

Ibídem, p. 9.

[80]

Ibídem, p. 12.

[81]

Ibídem, p. 12 y 13.

[82]

Ibídem, p. 13.

[83]

Ibídem, p. 13.

[84]

Ibídem, p. 52.

[85]

Ibídem, p. 55.

[86]

Ibídem, p. 55.

[87]

Algorithmic Accountability Act of 2019, disponible en <https://www.congress.gov/bill/116th-congress/house-bill/2231/all-info>, accedido el 15/1/2021.

[88]

Chae Yoon, "U.S. AI Regulation Guide: Legislative Overview and Practical Considerations", en RAIL The Journals of Robotics, Artificial Intelligence and Law Vol. 3, N° 1, (Jan-Feb 2020), Full Court Press, Washington, p. 21.

[89]

Chae, Yoon, U.S. AI..., p. 22.

[90]

Ídem.

[91]

Algorithmic Accountability Act of 2019, disponible en <https://www.congress.gov/bill/116th-congress/house-bill/2231/all-info>, accedido el 15/1/2021.

[92]

MacCarthy, Marck, An examination of the Algorithmic Accountability Act, 2019, p. 1, disponible en https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3615731 (Consultado el 29/06/2021).

[93]

Chae, Yoon, U.S. AI..., p. 22.

[94]

MacCarthy, Marck, An examination ..., p. 2.

[95]

Chae, Yoon, U.S. Al..., p. 22.

[96]

New, Joshua, "How to fix de Algorithmic Accountability Act", Center For Data Innovation, <https://datainnovation.org/2019/09/how-to-fix-the-algorithmic-accountability-act/> (Consultado el 29/06/2021).

[97]

Algorithmic Accountability Act of 2019, <https://www.congress.gov/bill/116th-congress/house-bill/2231/all-info> (Consultado el 29/06/2021).

[98]

MacCarthy, Marck, An examination ..., p. 2.

[99]

New, Joshua, "How to fix de Algorithmic Accountability Act", Center For Data Innovation, <https://datainnovation.org/2019/09/how-to-fix-the-algorithmic-accountability-act/> (Consultado el 29/06/2021).

[100]

Ídem.

[101]

Corte Suprema de Justicia de la Nación Argentina, (16/11/2004), "Hooft, Pedro C. F. c/ Provincia de Buenos Aires", Fallos, 327:5118.

[102]

Sánchez Caparrós, Mariana, Categorías Sospechosas, Astrea, Bs. As., 2020, p. 37 y sig.

[103]

Cfr. art. 1, CADH.

[104]

Ver en Sánchez Caparrós Mariana, ob. cit., p. 37.

[105]

Atala Riffo y Niñas vs. Chile, CIDH, San José, Costa Rica, 24/02/2012, Rubinzal Online, www.rubinzalonline.com.ar, RC J 4750/16.

[106]

Norín Catrimán, Segundo Aniceto y otros vs. República de Chile, CIDH, San José, Costa Rica, 29/05/2014; Rubinzal Online, www.rubinzalonline.com.ar, RC J 5531/19.

[107]

Sánchez Caparrós, Mariana, ob. cit., p. 38/39.

[108]

Ibídem, p. 39.

[109]

Norín Catrimán, Segundo Aniceto y otros vs. República de Chile, CIDH, San José, Costa Rica, 29/05/2014; Rubinzal Online, www.rubinzalonline.com.ar, RC J 5531/19.

[110]

Norín Catrimán, Segundo Aniceto y otros vs. República de Chile, CIDH, San José, Costa Rica, 29/05/2014, Rubinzal Online, www.rubinzalonline.com.ar, RC J 5531/19. Duque, Ángel Alberto vs. República de Colombia, CIDH, San José, Costa Rica; 26/02/2016, Rubinzal Online, RC J 5566/19.

[111]

Granier y otros (Radio Caracas Televisión) vs. Venezuela, CIDH, San José, Costa Rica, 22/06/2015, Rubinzal Online, www.rubinzalonline.com.ar, RC J 5884/15.

[112]

Corte IDH, (20/10/2016), "Trabajadores de la Hacienda Brasil Verde vs. Brasil", https://www.corteidh.or.cr/docs/casos/articulos/seriec_318_esp.pdf.

[113]

Gonzales Lluy, Talía Gabriela vs. Ecuador, CIDH, San José, Costa Rica; 01/09/2015, Rubinzal Online, www.rubinzalonline.com.ar, RC J 8819/19.

[114]

Sánchez Caparrós, Mariana, pág. 77 y sig.

[115]

Ídem.

[116]

Sánchez Caparrós, Mariana, ob. cit., p. 77 y sig.

[117]

Ampliar en Sánchez Caparrós, Mariana, ob. cit.

[118]

Ibídem, p. 11/12 y 77/79.

[119]

Ibídem, p. 36 y sig.

[120]

Ampliar en Sánchez Caparrós, Mariana, ob. cit.

[121]

Ídem.

[122]

Ibídem, p. 39.

[123]

Amunátegui Perelló, Carlos y Madrid Raúl, "Sesgo e Inferencia en Redes Neuronales ante el Derecho", en C. Aguerre, (Ed.). *Inteligencia Artificial en América Latina y el Caribe. Ética, Gobernanza y Políticas*, Bs. As., CETyS Universidad de San Andrés, 2020, p. 5, <https://guia.ai/documentos-2020/> (Consultado el 29/06/2021).

[124]

Asamblea General, ONU, Resolución 73/348, p. 15, <http://undocs.org/es/A/73/348> (Consultado el 29/06/2021).

[125]

Sánchez Caparrós Mariana, "Inteligencia Artificial, Sesgos y Categorías Sospechosas. Prevenir y mitigar la discriminación algorítmica", en *Tratado de Inteligencia Artificial y Derecho*, Tomo I, L.L., Buenos Aires, 2021, p. 315 y siguientes.

[126]

Corvalán, Juan G., *Inteligencia Artificial y Derechos Humanos*, Parte II, "Diario Constitucional y Derechos Humanos" Nro. 157, 2017, p. 2/3, <https://dpicuantico.com/sitio/wp-content/uploads/2017/07/Juan-Gustavo-Corvalan-Constitucional-10.07.2017.pdf>.

[127]

Ídem.

[128]

Sánchez Caparrós Mariana, "Inteligencia Artificial, Sesgos y Categorías Sospechosas. Prevenir y mitigar la discriminación algorítmica", en *Tratado de Inteligencia Artificial y Derecho*, Tomo I, L.L., Buenos Aires, 2021, p. 315 y siguientes.

[129]

El principio de explicabilidad algorítmica se refiere a la capacidad de explicar los procesos técnicos de un sistema de IA, así como las decisiones adoptadas con base en él. Ver Comisión Europea, Grupo de Expertos de alto nivel sobre inteligencia artificial, *Directrices para una IA fiable*, (abril, 2018), p. 24, <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es/format-PDF> (Consultado el 29/06/2021).

[130]

Ídem.

[131]

Ídem.

[132]

Parlamento Europeo, El impacto del Reglamento..., p. 14, https://www.europarl.europa.eu/stoa/en/document/EPRS_STU%282020%29641530 (Consultado el 29/06/2021).

[133]

Ídem.

[134]

Corvalán, Juan G., Perfiles..., p. 136. También, Corvalán, Juan G., "Inteligencia Artificial, Automatización...", p. 18. En ocasiones se puede efectuar un control adicional a través de un examen crítico de su funcionamiento, esto es, verificando sistemáticamente cómo cambia el resultado o salida si se modifica el valor de ciertas características de entrada, dejando las restantes características sin cambios, lo que permitirá comprender cuáles son las características que determinan cierta salida del sistema. Ver Parlamento Europeo, El impacto del Reglamento, p. 15, https://www.europarl.europa.eu/stoa/en/document/EPRS_STU%282020%29641530 (Consultado el 29/06/2021).

[135]

Gascón Marcer, Ana, "Derechos Humanos e Inteligencia Artificial", en Pérez Mirás A., Tarufel Lozano, Germán M., Raffiotta, Eduardo, Iadidicco María Pía (Dir.), Setenta Años de Constitución Italiana y Cuarenta Años de Constitución Española, Vol. V, Centro de Estudios Políticos y Constitucionales, 2020, p. 337/338, https://www.researchgate.net/publication/339687454_Derechos_Humanos_e_Inteligencia_Artificial (Consultado el 29/06/2021).

[136]

Ampliar en Corvalán, Juan G., Perfiles Digitales..., p. 136.

[137]

Cfr. art. 1, CADH.

[138]

En todos los casos, de aplicarse la doctrina se debería corroborar la presencia de una categoría sospechosa en los términos apuntados en el apartado II.2, de modo tal de que el empleo de la teoría se mantenga siempre en los límites de su propio cauce. Ampliar en Sánchez Caparrós, Mariana, Categorías Sospechosas, Astrea, Bs. As., 2020.

[139]

Ibídem, p. 39.

[140]

Sánchez Caparrós Mariana, "Inteligencia Artificial, Sesgos y Categorías Sospechosas. Prevenir y mitigar la

discriminación algorítmica", en Tratado de Inteligencia Artificial y Derecho, Tomo I, L.L., Buenos Aires, 2021, p. 315 y siguientes.

[141]

Corvalán, Juan G., Inteligencia Artificial..., p. 2/3, <https://dpicuantico.com/sitio/wp-content/uploads/2017/07/Juan-Gustavo-Corvalan-Constitucional-10.07.2017.pdf>.

[142]

Comisión Europea, Grupo de Expertos de alto nivel sobre inteligencia artificial, Directrices..., p. 49, <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es/format-PDF>. Ver también Asamblea General, ONU, Resolución 73/348, p. 10, <http://undocs.org/es/A/73/348> (Consultados el 29/06/2021).

[143]

Comisión Europea, Grupo de Expertos de alto nivel sobre inteligencia artificial, Directrices..., p. 51, <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es/format-PDF> (Consultados el 29/06/2021).

[144]

Ídem, p. 7.

[145]

Asamblea General, ONU, Resolución 73/348, p. 4, <http://undocs.org/es/A/73/348> (Consultados el 29/06/2021).

[146]

Comisión Europea, Directrices, p. 1, <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es/format-PDF>.

[147]

Corvalán, Juan G., Inteligencia Artificial..., p. 2/3, <https://dpicuantico.com/sitio/wp-content/uploads/2017/07/Juan-Gustavo-Corvalan-Constitucional-10.07.2017.pdf>.

[148]

Sánchez Caparrós Mariana, "Inteligencia Artificial, Sesgos y Categorías Sospechosas. Prevenir y mitigar la discriminación algorítmica", en Tratado de Inteligencia Artificial y Derecho, Tomo I, L.L., Buenos Aires, 2021, p. 319 y siguientes.

[149]

Corvalán, Juan G., Inteligencia Artificial..., p. 2/3, <https://dpicuantico.com/sitio/wp-content/uploads/2017/07/Juan-Gustavo-Corvalan-Constitucional-10.07.2017.pdf> (Consultados el 29/06/2021).

[150]

Parlamento Europeo, El impacto..., p. 21, https://www.europarl.europa.eu/stoa/en/document/EPRS_STU%282020%29641530 (Consultados el 29/06/2021).

[151]

Sánchez Caparrós Mariana, "Inteligencia Artificial, Sesgos y Categorías Sospechosas. Prevenir y mitigar la discriminación algorítmica", en Tratado de Inteligencia Artificial y Derecho, Tomo I, La Ley, Buenos Aires, 2021, p. 319 y siguientes.

[152]

Allen Robin QC y Masters Dee, "In the matter of automated data processing in government decision making", (7 September 2019), open opinion for The Legal Education Foundation, p. 46, <https://www.cloisters.com/wp-content/uploads/2019/10/Open-opinion-pdf-version-1.pdf>.

[153]

Ibídem, p. 39.

[154]

Sánchez Caparrós Mariana, "Inteligencia Artificial, Sesgos y Categorías Sospechosas. Prevenir y mitigar la discriminación algorítmica", en Tratado de Inteligencia Artificial y Derecho, Tomo I, La Ley, Buenos Aires, 2021, p. 319 y siguientes.