ANTISUBORDINACIÓN Y DISCRIMINACIÓN ALGORÍTMICA
ANTI-SUBORDINATION AND ALGORITHMIC DISCRIMINATION

Anna Capellà i Ricart
Investigadora Postdoctoral
Institut de Dret i Tecnologia
Universitat Autònoma de Barcelona

RESUMEN

La perspectiva antisubordinatoria del derecho antidiscriminatorio se basa en la noción de igualdad como no exclusión. Tal como opera actualmente en distintos contextos, la inteligencia artificial contribuye a la perpetuación de sistemas sociales de opresión/subordinación que generan exclusión. Este artículo analiza la dicotomía entre las corrientes anticlasificación y antisubordinación y el impacto de ambas perspectivas en el combate contra la discriminación algorítmica. A tal efecto, se estudian cuatro elementos clave para entender la perspectiva antisubordinatoria (los estereotipos, la no neutralidad, la descontextualización y la monocultura) que muestran que la misma crítica que se hace desde la perspectiva antisubordinatoria al Derecho puede dirigirse a la inteligencia artificial, dado que en estos dos ámbitos se verifica una concentración de poder. Finalmente, proponemos algunas medidas antisubordinatorias orientadas a hacer frente a la discriminación algorítmica.

PALABRAS CLAVE

Antisubordinación, inteligencia artificial, Derecho, opresión, algoritmo.

ABSTRACT

The anti-subordination perspective of anti-discrimination law is based on the notion of equality as non-exclusion. As it currently operates in different contexts, artificial intelligence contributes to the perpetuation of social systems of oppression/subordination that cause exclusion. This article analyzes the dichotomy between the anti-classification and anti-subordination perspectives and their impact in the fight against algorithmic discrimination. To this end, four key elements are studied to understand the anti-subordination perspective (stereotypes, non-neutrality, decontextualization and monoculture) that show that the same criticism made from the anti-subordination perspective to the Law can be directed to artificial intelligence (as both are areas in which power is concentrated). Finally, we propose some anti-subordination measures aimed at confronting algorithmic discrimination.

KEYWORDS

Anti-subordination, artificial intelligence, Law, oppression, algorithm.

DOI: https://doi.org/10.36151/TD.2024.112

ANTISUBORDINACIÓN Y DISCRIMINACIÓN ALGORÍTMICA

Anna Capellà i Ricart
Investigadora Postdoctoral
Institut de Dret i Tecnologia
Universitat Autònoma de Barcelona

1. INTRODUCCIÓN

El análisis masivo de datos —y, específicamente, la elaboración de perfiles— se basa en el descubrimiento de correlaciones entre datos alojados en bases y su finalidad es identificar, representar y clasificar a personas como miembros de grupos o categorías. A cada persona clasificada en un grupo o una categoría se le puede aplicar un perfil de grupo (es decir, ciertas características que el algoritmo considera probable que sean comunes en las personas que lo integran). Dado que la disponibilidad de datos relevantes crece de manera exponencial, cuanto más sofisticado sea el perfil del grupo, tanto más se inclinará el algoritmo a modelar un perfil personalizado y más sutilmente discriminará entre las personas que forman parte del grupo y las que no pertenecen a él (Schreurs *et al.,* 2008: 242).

Si el análisis masivo de datos es el conjunto de técnicas y tecnologías que permiten explorar y explotar grandes cantidades de datos de forma automática y semiautomática con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los sujetos en un determinado contexto (Rivas Vallejo, 2020: 103), siempre que los datos estén ligados a comportamientos sociales (basados el estructuras de opresión, subordinación y exclusión) los resultados que se obtengan a partir de ellos serán discriminatorios. Las tareas que realizan los sistemas algorítmicos suponen la aplicación de estos patrones repetitivos, tendencias o reglas a conjuntos de datos para extraer resultados.

En la medida en que estos patrones, tendencias o reglas tomen como referencia contextos sociales determinados, los sistemas algorítmicos tienderán a reproducirlos. Por lo tanto, es muy probable que el uso de sistemas algorítmicos predictivos en un contexto social determinado implique la obtención de resultados discriminatorios, dado que los resultados simplemente reflejarán una sociedad desigual y discriminatoria que oprime a determinados grupos de población que comparten ciertos atributos protegidos.

Para hacer frente a la discriminación que genera la utilización de sistemas algorítmicos, se pretende «reparar» estos sistemas mediante métricas predefinidas de «equidad algorítmica», término técnico utilizado en la comunidad de aprendizaje automático para describir investigaciones y aplicaciones que tratan de abordar la discriminación en sistemas algorítmicos (West, 2020: 4).

En este trabajo ponemos en cuestión la capacidad de las herramientas de equidad algorítmica presentadas hasta ahora —mayoritariamente basadas en la equidad formal— para hacer frente a la discriminación. En la sección 2 presentamos la perspectiva anticlasificatoria (o formal) del Derecho antidiscriminatorio y la perspectiva antisubordinatoria (o sustantiva) del Derecho antidiscriminatorio. Seguidamente, la (sección 3), analizamos cuatro elementos para entender la perspectiva antisubordinatoria: los estereotipos, la no neutralidad, la descontextualización y la monocultura; del estudio de cada uno de estos elementos extraemos la crítica que la perspectiva antisubordinatoria ha formulado al abordaje de la discriminación por parte del Derecho y la extrapolamos al tratamiento de la discriminación generada por la inteligencia artificial (en adelante, IA). En la sección 4 proponemos algunas medidas orientadas a entender la discriminación producida por los sistemas de la IA desde la perspectiva antisubordinatoria. Finalmente, en la sección 5 presentamos unas reflexiones conclusivas.

2. ANTICLASIFICACIÓN Y ANTISUBORDINACIÓN

La cláusula específica de no discriminación incorporada a los textos constitucionales y a los instrumentos internacionales de derechos humanos pretende proteger a determinados grupos o colectivos que, debido a la minusvaloración, la subordinación y la postergación social que han padecido a lo largo de la historia, no han participado, de hecho, en los procesos de formación de las normas (Añón Roig, 2013: 131). Sin embargo, la interpretación dominante del principio de igualdad en el en el campo del Derecho reenvía a la idea de «igualdad de todos», es decir, a la atribución del mismo estatus jurídico a todas las personas, y no reconoce la existencia de grupos subordinados. Semejante aproximación a la igualdad jurídica y política está basada en un falso universalismo, dado que en el ámbito jurídico la abstracción jurídica del «individuo» o «sujeto» del que se predica la igualdad ha sido edificada y desarrollada tomando como referencia un tipo *específico* de sujeto (hombre, blanco, propietario…) (Barrère Unzueta, 2008: 54). El principio normativo de acuerdo con el cual todas las personas tienen el mismo estatus jurídico (y, por tanto, deben ser consideradas

iguales por el Derecho) presume la neutralidad del Estado y, por ello, no considera las diferencias entre las personas en la ley ni en la aplicación de la ley. Desde el momento en que las personas ya tienen reconocidos los mismos derechos, los mecanismos de protección previstos en el ordenamiento jurídico se limitan a evitar que el ejercicio abusivo (pero individualizado) de los derechos de unas no interfiera en los derechos (individuales) de otras. En definitiva, la idea de justicia se traduce en la igualdad del estatus jurídico individual (o, si se prefiere, en la indiferenciación tanto en la ley como en la aplicación formal de la ley) (Barrère Unzueta, 2008: 54-55).

La función del Derecho antidiscriminatorio se ha interpretado desde dos perspectivas distintas. Por un lado, la corriente anticlasificación sostiene que el Derecho antidiscriminatorio debe garantizar la igualdad formal entre los individuos, considerando que cualquier trato desigual solo resulta admisible si es razozable y no arbitrario poner la atención en que cualquier trato es igual cuando es razonable y no arbitrario, rechazando cualquier trato diferenciado entre personas que no guarde relación funcional con la finalidad perseguida. Desde este enfoque, la discriminación se asocia como ruptura de la igualdad de trato en términos básicamente individuales, focalizándose en la igualdad formal y asumiendo una visión litigante (Añón Roig, 2022: 41).

Por ejemplo, desde el punto de vista de la privacidad, la anticlasificación prohíbe cualquier diferenciación basada en un rasgo o motivo protegido. El razonamiento que subyace a esta concepción es que, si puede prevenirse el acceso a cierta información especialmente protegida, en el futuro se evitará la discriminación basada esta información protegida (Roberts, 2015: 2123-2124). Sin embargo, algunos autores han puesto de manifiesto que la perspectiva anticlasificación puede no funcionar en los sistemas algorítmicos (Barocas *et al.,* 2019: 199). Para evitar la clasificación de las personas a partir de su pertenencia a uno o varios grupos en los procesos de tratamiento masivo de dados, el modelo algorítmico debería ignorar o suprimir intencionadamente todos los atributos protegidos y todos los factores correlacionados con estos. Esta acción generaría un modelo con resultados mucho menos precisos (Barocas y Selbst, 2016: 727). Además, cabe poner en cuestión la posibilidad de desarrollar este enfoque, dado que en muchas ocasiones no se puede determinar qué datos correlacionados acabarán suministrando información que la legislación considera protegida.

Desde un punto de vista crítico, se ha señalado que la perspectiva anticlasificación ignora la historia de minusvaloración y opresión que han sufrido ciertos grupos de población (Barrère Unzueta, 2018: 13). Esta aproximación «ciega» es la que se sirve de base para hacer frente a la discriminación algorítmica. Como señala Green (2022: 4), la lucha contra la discriminación en el ámbito de los sistemas algorítmicos restringe el margen de apreciación de la información a momentos decisorios concretos —datos de entrada y salida de la decisión en cuestión—, es decir, desvincula los datos del contexto en el que han sido extraídos y se limita a asegurar que el resultado del tratamiento no sea formalmente discriminatorio, pero no tiene en cuenta por qué razón los datos son unos y no otros; esta desatención, asociada a la definición de la equidad como un mero atributo técnico de los algoritmos, permite ignorar abiertamente la historia de minusvaloración y opresión que han sufrido ciertos grupos

de población (Green, 2022: 12). Por su parte, West (2020: 59) observa que, más allá de que los discursos son cada vez más interdisciplinares, para hacer frente a la discriminación algorítmica se ha consolidado un corriente dominante centrada en los marcos computacionales del problema y en el desarrollo de herramientas diseñadas para identificar y mitigar sesgos en los sistemas algorítmicos —por ejemplo, corregir el sesgo estadístico en conjuntos de datos para que se alcancen tasas de precisión en todas las categorías demográficas—. Este enfoque dominante parte de la premisa de que hay problemas en el funcionamiento de los sistemas de IA que pueden ser solucionados e incluso regulados con el objetivo de legitimar la venta de los bienes y servicios de la IA, pero omite señalar que este tipo de problemas de funcionamiento viene causado por un conflicto estructural y dinámico que no puede solventarse tecnológicamente. Como acabamos de sugerir, el reconocimiento de este déficit podría dificultar la comercialización de bienes y servicios de IA para su utilización en contextos donde pueden producirse situaciones de discriminación. Adicionalmente, este planteamiento se aleja de la concepción según la cual, si bien los algoritmos pueden remediar problemas sociales, en muchos casos son herramientas innecesarias o incluso perjudiciales (Green, 2022: 18).

Por su parte, la corriente antisubordinación sostiene que la principal finalidad del Derecho antidiscriminatorio es hacer frente a la exclusión social, la opresión y la subordinación de ciertos grupos de población y que, en este sentido, debe articularse en torno a la noción de igualdad entendida como no sometimiento o no exclusión. Esta perspectiva considera especialmente relevante la existencia de personas y grupos de personas que, al haber sido histórica y sistemáticamente excluidas de distintos ámbitos de la vida social, y haberse visto subordinadas y oprimidas por determinadas estructuras sociales de poder y dominación, rompen el esquema basado en la presunta igualdad entre todas las personas (Añón Roig, 2022: 41 y 42), y propone expandir el ámbito de análisis a consideraciones relacionales y estructurales que trascienden la toma de decisiones en casos específicos (Green, 2022: 8 y 16). La ruptura de aquel esquema y la falta de atención a la dimensión estructural de la discriminación tiene implicaciones no menores, dado que los grupos subordinados ven limitadas sus oportunidades vitales y su acceso a los bienes y servicios básicos, y se encuentran, por ello, en una situación de desequilibrio en el desarrollo de sus relaciones sociales (Álvarez del Cuvillo, 2022: 189). La perspectiva antisubordinación centra su atención en la dimensión material de la igualdad y aboga por la eliminación de las desventajas que afectan los grupos oprimidos a través de la incorporación al ordenamiento jurídico de medidas de trato diferencial, la puesta en marcha de estrategias de protección de la diversidad y la implementación políticas que integren las particularidades de cada grupo en los instrumentos reguladores generales con el objetivo de reconducir y subsanar la historia de subordinación que han padecido aquellos grupos (Soriano Arnanz, 2021: 10 y 11). Se trata, pues, de un planteamiento crítico informado por una finalidad transformadora de las estructuras que reproducen la desigualdad y la discriminación cuyo arsenal argumentativo se orienta a poner en evidencia la insuficiencia y aun el agotamiento del modelo formalista/anticlasificatorio para abordar la discriminación (Morondo Taramundi, 2023: 713 y 714). La clave del estudio de la perspectiva antisubordinatoria radica en el análisis del modo en que las sociedades dan respuesta a las diferencias en los atributos y capacidades de las personas

pertenecientes a determinados colectivos (por ejemplo, el color de la piel, el sexo o la capacidad física), y su objetivo último es luchar para que todas sean tratadas con igual respeto con independencia de aquellos rasgos diferenciadores (Green, 2022: 8).ç

Esta concepción, el Derecho antidiscriminatorio no puede prescindir de la crítica a determinados planteamientos sustentados en una concepción del poder que se ha denominado grupal, sistémica y estructural, dado que, como hemos dicho, responde a la idea de opresión o «poder sobre» personas excluidas o subordinadas que, en su condición de tales, adquieren conciencia e identidad política (Barrère, 2018: 20). La perspectiva antisubordinación entiende, por tanto, que la discriminación no solo resulta perjudicial cuando afecta a una persona específica, sino también (o, sobre todo) cuando el sujeto pasivo es todo el colectivo social (o los colectivos sociales) a los que aquella pertenece, puesto que en tal caso el bien lesionado es la posición del colectivo en la comunidad (Barocas *et al.*, 2019: 79). Es otros términos, si bien los actos discriminatorios concretos son individualizables, en realidad solo son una consecuencia de opresiones sociales subyacentes (Añón Roig, 2021: 36). En el proceso de individualización —que inevitablemente comporta la ignorancia o el desprecio de la importancia del colectivo— la discriminación es abordada como un problema de las personas excluidas y no como un ejemplo que ilustra la existencia de opresiones no abordadas (Bodelón, 2010: 88). Los mecanismos reparadores centrados en situaciones concretas de discriminación perpetradas contra individuos fallan cuando pretenden asegurar la igualdad de tratamiento de los propios individuos en riesgo de discriminación sistémica, y también fracasan a la hora de proteger a otras personas que pueden sufrir daños colaterales derivados de las estructuras sociales que discriminan, de modo que están mal equipados para hacer frente al daño masivo y las prácticas sistémicas ilegales (Farkas, 2014: 26). El rol fundamental que corresponde al grupo en el Derecho antidiscriminatorio interpela a las estructuras jurídicas basadas en una concepción de la igualdad de corte liberal y basada en el individuo y pone de manifiesto sus limitaciones (Barrere Unzueta, 2001: 146; Morondo Taramundi, 2023: 715).

El enfoque de la perspectiva antisubordinación está orientado a identificar el modo en que los sistemas jurídicos plasman las desigualdades que tienen su razón de ser en las estructuras de poder capaces de ordenar las relaciones sociales para atribuir o asignar estatus (subordinados o inferiores, privilegiados o superiores) y, por tanto, para robustecer las dinámicas e inercias que reproducen estas relaciones de subordinación (Barrère, 2018: 32; Green, 2022: 9 y 10). Barocas *et al.,* afirman que la ley es un instrumento que puede emplearse tanto para crear las condiciones que propicien el florecimiento de la discriminación como para contrarrestarla. En este sentido, el Derecho es (puede ser) tanto una herramienta de subyugación como una herramienta de liberación (Barocas *et al.*, 2019: 143). Esta disyuntiva es aplicable a la inteligencia artificial: en efecto, la utilización de la IA puede perpetuar la discriminación, pero también puede contribuir a luchar contra ella. Ahora bien, lo relevante no son los usos (buenos) que hipotéticamente se le puedan dar tanto a los sistemas jurídicos como a la IA (que, por cierto, en el caso de la AI legitiman su utilización como herramientas de diferenciación para la igualdad con vistas a materializar un cambio altamente improbable que solo se producirá a través de transformaciones socia-

les mucho más profundas). La cuestión central es quién detenta el poder y cómo se están empleando actualmente estos instrumentos: la utilización de la IA en diversos contextos sociales evidencia una vocación clara de perpetuación del *statu quo* opresivo que aqueja a ciertos grupos de población.

A este respecto, West plantea que, desde sus inicios, la cultura de la IA ha estado dominada por los hombres, y que, al moldear el comportamiento relacional a través de las máquinas, estos tomaron como referencia una concepción particular de la inteligencia humana basada en la masculinidad adulta, blanca y de clase alta, de modo que la discriminación está se integró en el campo de la IA en el mismo momento de su fundación. A título ilustrativo, cabe hacer referencia al uso recurrente de personajes femeninos virtuales en los sistemas de asistencia telemática que replican los estereotipos de género e imprimen a su ejecutoria la docilidad y la disposición servil tradicionalmente atribuida a las mujeres (West, 2020: 9). En otro orden de cosas, la IA también promueve la servidumbre en lugar de la autonomía y la independencia.

En la actualidad, el Derecho antidiscriminatorio no se ha comprometido de manera clara con el principio de antisubordinación (y, por consiguiente, este principio tampoco se ha trasladado al ámbito de la toma de decisiones a través sistemas algorítmicos). Por ello, sigue dominando la perspectiva que concibe el principio de igualdad de trato como indiferenciación o como trato razonable no arbitrario (Añón Roig, 2022: 43-44).

3. ANTISUBORDINACIÓN E IA. ¿QUÉ ELEMENTOS HAY QUE TENER EN CUENTA?

En esta sección analizaremos el impacto de los estereotipos (3.1), la presunción de neutralidad (3.2), la descontextualización (3.3) y la promoción de una monocultura (3.4) en la discriminación derivada del uso de sistemas algorítmicos en contextos sociales. Además, pondremos en relación el impacto de estos elementos en los resultados que presenta la IA con la crítica al modo en que el Derecho antidiscriminatorio aborda habitualmente esta cuestión.

3.1. ESTEREOTIPOS

Un elemento clave para entender la discriminación son los estereotipos. Los estereotipos son visiones generalizadoras o preconcepciones de los atributos, características o roles que, se supone, deben cumplir los miembros de un colectivo social concreto y que contribuyen a reforzar la presunción de que cualquier persona que se integre en el colectivo social actuará de conformidad con aquellas visiones generales o preconcepciones (por ejemplo, las mujeres son cuidadoras por naturaleza) (Cook y Cusack, 2010: 11). Se trata de representaciones mentales (mecanismos cognitivos neutros normalmente de carácter inmutable) que nos ayudan a reducir la complejidad de la realidad y son interseccionales,

es decir, están constituidos simultáneamente por diferentes ejes, entre ellos el género, la raza, la clase social o la orientación sexual (Ghidoni y Morondo Taramundi, 2022: 62). A pesar de que se consideran neutros, su función, su uso o su explotación por parte de los sistemas de opresión puede tener impactos positivos o negativos. Esta caracterización de los estereotipos permite analizarlos como elementos sistémicos y difusos, así como alejar el análisis de la discriminación de algunos elementos clave, entre ellos los efectos que producen (como preeminentemente sucede en los casos de discriminación algorítmica) (Ghidoni y Morondo Taramundi, 2022: 41-43).

Los estereotipos forman parte del proceso de subordinación, que construye las identidades y el estatus de los grupos desaventajados y dota de significado al comportamiento social sin que exista ningún mecanismo especifico que aísle su influencia del razonamiento jurídico (Morondo Taramundi, 2023: 710, 718 y 726) o, en relación con la cuestión que nos ocupa, del desarrollo de la IA. Los estereotipos tienen un origen social (están estructurados socialmente como convenciones). Los sistemas jurídicos los absorben y los refuerzan a través de la producción, la interpretación y la aplicación del Derecho (Añón Roig, 2021: 45), y justifican, naturalizan e invisibilizan las estructuras de poder (por ejemplo, en las normas que prevén los permisos para madres –más largos y obligatorios– y para los padres —optativos y más cortos—) (Ghidoni y Morondo Taramundi, 2022: 46 y 49). Los estereotipos desempeñan un rol fundamental en la perpetuación de la discriminación y son difíciles de identificar y neutralizar a través de las categorías jurídicas (Morondo Taramundi, 2023: 722).

Los estereotipos tienen un impacto sobre la persona —dado que presumen en ella ciertos atributos, características y roles que la despersonalizan, impiden su libre individualización e influyen en la construcción de su identidad—, pero también sobre el grupo (o los grupos) al cual pertenece —pues los rasgos o roles atribuidos por el estereotipo al colectivo (ya sean positivos o negativos) determinarán su posición en la jerarquía social: a este respecto la cuestión clave es quién generaliza o puede generalizar y para qué generaliza— (Ghidoni y Morondo Taramundi, 2022: 57-60).

Añón Roig subraya que los estereotipos perpetúan relaciones de poder opresivas, excluyen a algunas personas del acceso a bienes básicos, y reducen la autonomía y el autorrespeto porque repercuten en dimensiones que afectan centralmente al control sobre la propia vida o a las decisiones básicas relacionadas con ella (Añón Roig, 2021: 48).

Centrándonos en la IA, cabe apuntar que, a pesar de que se ha presentado como una tecnología potente, omnipresente y casi infalible para resolver problemas (decisionales y de eficiencia), en esencia se limita a encontrar patrones en enormes cantidades de datos (Powles y Nissenbaum, 2018). Si bien algunos patrones que se encuentran en los datos de entrenamiento representan conocimientos adquiridos (fumar está asociado al cáncer), otros patrones representan estereotipos que operan socialmente (a las niñas les gusta el rosa y a los niños les gusta el azul). Los algoritmos de aprendizaje automático no disponen de herramientas capaces de distinguir estos dos tipos de patrones, de manera que extraerán estereotipos de la misma manera en que extraen conocimiento (Barocas *et al.,* 2019: 10). Según Ghidoni y Morondo Taramundi, los «sesgos sin prejuicio» de las máquinas han lle-

vado también a replantearse cómo ingresan y cómo operan los estereotipos en los procesos de decisión en general (Ghidoni y Morondo Taramundi, 2022: 40).

Los estereotipos también están presentes en la preparación previa para que el sistema de aprendizaje automático pueda llegar a detectar patrones. Por ejemplo, el etiquetaje de datos puede plantear conflictos cuando las personas que llevan a cabo esta tarea deben realizar elecciones difíciles para determinar qué etiquetes son más adecuadas (es decir, cada ejemplo puede cumplir ciertos criterios, pero no todos, para ser asignado a una etiqueta concreta). También puede ocurrir que el etiquetaje no sea lo suficientemente preciso como para capturar diferencias significativas entre los casos, circunstancia que motiva que las personas que hacen el etiquetaje realicen juicios de valor que influyen en la toma de decisiones posterior (Barocas y Selbst, 2016: 681). Un caso representativo puede ser el de ImageNet, una base de datos que funciona con la sistemática de clasificación de imágenes según nueve categorías (que después se desglosan en subcategorías): planta, formación geológica, objeto natural, deporte, artefacto, hongo, persona, animal y variado. A pesar de que algunos conceptos —por ejemplo, 'manzana'— puedan ser fáciles de catalogar, hay otros que plantean una mayor complejidad. Cuando nos referimos a «persona» se integran etiquetas como «persona sin habilidades» o «mulato» que integran imágenes permeadas por estereotipos de raza, genero, edad y habilidad. No hay, pues, categorías neutras en ImageNet, dado que la selección de las imágenes siempre interactúa con el significado de las palabras (Crawford, 2021: 138-147). Los creadores de ImageNet recurrieron a la plataforma Mechanical Turk (MTurk), de Amazon, un mercado laboral en línea que permite a los individuos y las corporaciones contratar trabajadores bajo demanda para realizar tareas simples (Barocas *et al.*, 2019: 238). A los trabajadores de MTurk se les daba la tarea de clasificar cincuenta imágenes (extraídas de navegadores web) por minuto en las categorías propuestas. En los resultados recogidos en la base de datos pueden encontrarse estereotipos, errores y contrasentidos absurdos, entre ellos que una persona que se encuentra en la playa sea etiquetada como «cleptómana» o que una persona adolescente que lleva una sudadera deportiva sea clasificada como «perdedora». A pesar de que se intenten suprimir o matizar los peores ejemplos, la perspectiva esta fundamentalmente construida con base en la separación entre los datos, las personas, los sitios y los contextos de donde provienen a fin de transmitir una visión del mundo técnica que pretende infundir una suerte de objetividad a materiales culturales complejos y de procedencia heterogénea (Crawford, 2021: 138-147).

También se ha alertado de que los sistemas de búsqueda y recomendación satisfacen las necesidades de algunos usuarios mejor que las de otros porque los perfiles de consumidores de los primeros son similares a los de las personas que han creado los sistemas. Las personas que crean los sistemas pueden privilegiar ciertos contenidos sobre otros y generar daños representacionales mediante la amplificación y la perpetuación de estereotipos culturales, dado que influyen en la selección de los mensajes más difundidos (Barocas *et al.*, 2019: 20 y 192). Adicionalmente, se ha observado que el lenguaje y las imágenes capturan tal variedad de estereotipos culturales que los métodos de eliminación de los sesgos algorítmicos no suprimen, en realidad, el sesgo, sino que simplemente lo esconden (Gonen y Goldberg, 2019: 5).

Los sistemas sociales opresivos (machismo, racismo, clasismo) generan estereotipos opresivos (por ejemplo, la madre abnegada, el inmigrante gorrón, el pobre holgazán), de manera que los estereotipos deben ser entendidos como evidencias de los mecanismos que pretenden invisibilizar (Ghidoni y Morondo Taramundi, 2022: 54). En este sentido, el hecho que los sistemas de IA aplicados a contextos sociales simplifiquen la realidad, la presenten desde el punto de vista de la cultura mayoritaria (cfr. apartado 3.4) y amplifiquen sus resultados puede facilitar la detección de los estereotipos que pretenden reforzar y perpetuar los sistemas de opresión (Wachter, 2021: 372).

Sobre todo, hay que tener en cuenta que el diseño de métodos para hacer frente a la discriminación algorítmica sin tener en cuenta los estereotipos que tiñen la sociedad no solo implica la inefectividad de aquellos dispositivos, sino también la participación de la IA en la perpetuación de la discriminación. En la construcción del concepto de «discriminación» ya se ha recorrido este camino, que ha llevado a fomentar la confusión sobre su significado y a propiciar que personas que no han sufrido una historia de opresión puedan considerarse discriminadas. Por ejemplo, como plantea Lousada Arochena, las leyes de igualdad de género clásicas se caracterizan (entre otras cosas) por el hecho de que se han construido tomando únicamente en consideración el concepto de sexo, es decir, como si los estereotipos de género no existieran. El efecto colateral de este enfoque legislativo ha sido la extensión de la protección contra la discriminación a los dos sexos y la creación del espejismo del varón discriminado (Lousada Arochena, 2022: 4).

3.2. ¿NEUTRALIDAD?

En este apartado pretendemos mostrar que las críticas a la supuesta neutralidad del Derecho son las mismas que pueden formularse a la supuesta neutralidad de la IA.

En línea con la teoría crítica del Derecho, Lousada Arochena sostiene que el Derecho no representa la razón universal, sino la razón de los hombres en cuanto a detentadores del poder. De esta manera, la perspectiva social y cultural de los hombres es la que ha servido de base para construido la totalidad de un ordenamiento jurídico que pretende aparentar neutralidad (Lousada Arochena, 2022: 6 y 7). Uno de los objetivos del análisis antisubordinatorio es identifica las diversas vías a través de la que se plasman las desigualdades en los sistemas jurídicos y los modos en que estos ayudan a asignar estatus y establecer dinámicas e inercias que reproducen las relaciones de subordinación. En otros términos, uno de los objetivos del Derecho antisubordinatorio es poner en entredicho la supuesta neutralidad del derecho (Barrère, 2018: 32).

En un principio, las decisiones tomadas por sistemas de IA se consideraban más objetivas —y por tanto más neutrales— que las decisiones tomadas por personas. La razón principal de la atribución de una supuesta mayor neutralidad a los sistemas de IA no era otra que la visión encomiástica, apologética y acrítica de la tecnología: dado que los resultados que arrojan las máquinas arrojan derivan del uso de mecanismos técnicos, son más objetivas que las personas. Sin embargo, cuando intervienen en procesos decisorios que influyen en las personas, las máquinas tienen los mismos sesgos que estas (pues han sido creadas por

personas). Es precisamente la tecnofilia la que sigue amparando la supuesta neutralidad de la IA y dificulta que esta pueda ser evaluada por la población. Birhane *et al.* (2022: 178) publicaron un estudio orientado a analizar una serie de artículos influyentes en la disciplina del aprendizaje automático; los resultados mostraron que tan solo un 1 % de los encuestados mencionó o discutió sus posibles efectos negativos. Los autores concluyen que la tecnología no es una disciplina neutral en relación con los valores que promueve, sino que tiene una carga social y política, dado que fomenta la concentración de recursos, herramientas, conocimiento y poder en manos de actores que ya son poderosos, y prioriza y operacionaliza valores como el desempeño, la generalización, la eficiencia y la novedad (Birhane *et al.,* 2022: 182). Además, el tecnicismo también establece distancias entre las personas y permite que las reacciones habituales de indignación ante la discriminación se diluyan. En este sentido, Bigman *et al.* (2023: 4-6) afirman que la discriminación algorítmica causa menos indignación moral que la discriminación humana y sugieren que esta actitud puede estar motivada por el hecho de que las personas tienen menos predisposición a atribuir motivaciones negativas a un algoritmo —ya que las personas son percibidas como seres capaces de intención y antipatía y, por tanto, es más probable que tengan una motivación perjudicial, mientras que los algoritmos no son percibidos como capaces de intención y antipatía y, por tanto, es más improbable que se atribuya a su resultado una motivación perjudicial—.

Green (2021: 250-253 y 259) plantea que las personas que trabajan en el análisis masivo de datos deben reconocerse como actores políticos implicados en las construcciones normativas de la sociedad. En autor compara el impacto de los analistas de datos actuales con el de los ingenieros que organizaron la circulación cuando apareció el coche; a pesar de que estos últimos consideraban que su intervención era meramente tecnológica (establecer las señales de tráfico y los sistemas de temporización de los semáforos), en realidad diseñaron todo el entramado circulatorio para privilegiar la ubicuidad del automóvil, que desde ese momento se adueñó de las calles. En este sentido, el hecho de orientar el análisis masivo de datos hacia nuevas aplicaciones es fundamentalmente antidemocrático, dado que permite a las personas que trabajan en este campo (mayoritariamente del mismo grupo social) dar forma a la sociedad sin deliberación ni responsabilidad y pone de manifiesto la ilusoria creencia de que el desarrollo científico requiere pocas responsabilidades morales o políticas y que se lleva a cabo con neutralidad, un apriorismo que omite el hecho de que resulta imposible llevar a cabo ninguna tarea sin estar influenciado por determinados antecedentes, valores e intereses. Los esfuerzos encaminados a favorecer la neutralidad no conforman una opción política neutra, sino más bien una opción fundamentalmente conservadora que perpetúa el *statu quo* y refleja una actitud aquiescente con los valores sociales y políticos dominantes.

Green (2021: 254 y 259) señala críticamente que, si bien las personas que se dedican al análisis masivo de datos reconocen que este no pude proporcionar soluciones perfectas a problemas sociales, generalmente dan por descontado que contribuyen al «bien social» y que esa aplicación de la tecnología es una estrategia adecuada para el progreso social. Sin embargo, considerados desde una perspectiva de la igualdad sustantiva y antiopresión,

muchos esfuerzos de la ciencia de datos para hacer el bien no lo están haciendo de manera consistente. Sería, por tanto, necesario desarrollar una metodología rigurosa que relacionara las intervenciones algorítmicas con los impactos sociales (*ibid.*: 255), sobre todo respecto a sus impactos sobre colectivos sociales determinados, dado que muchos actores políticos y empresas de tecnología se benefician del cumplimiento de estándares en relación con la equidad algorítmica formal sin hacer planteamientos políticos o económicos significativos (Green, 2022: 23). Por ejemplo, no basta desarrollar algoritmos reconociendo que los datos sobre los delitos están sesgados; es necesario, además, reconocer que las definiciones de delito, el conjunto de instituciones encargadas de sancionarlo y las intervenciones que esas instituciones brindan son el resultado de procesos políticos históricos cargados de discriminación (Green, 2021: 258).

Por otra parte, Barocas *et al.* (2019: 24) apuntan que focalizar la atención en el aseguramiento de la equidad algorítmica formal supone eclipsar los debates sobre la legitimidad de las decisiones tomadas mediante sistemas algorítmicos. Los autores afirman que muchas empresas han adoptado el discurso de la equidad y que les resulta relativamente fácil garantizar la paridad en las decisiones entre grupos demográficos sin abordar las preocupaciones de legitimidad. Precisamente, la incorporación de la perspectiva antisubordinatoria a este análisis podría evidenciar que la toma de decisiones no es legítima (en cuanto justa) porque el sistema algorítmico participa en la perpetuación de la situación de privilegio de ciertos grupos sociales frente a otros. En este sentido, resultan especialmente relevantes los planteamientos que apuntan que las intervenciones orientadas a abordar los perjuicios causados por sistemas algorítmicos deberían centrarse en las instituciones subyacentes, dado que, en realidad, la adopción de decisiones mediante sistemas automatizados permite y aun legitima la inacción de las instituciones e impide el impulso de reformas estructurales necesarias (Barocas *et al.*, 2019: 218). Además, el planteamiento anclado en la pertinencia de la equidad algorítmica formal implica la difusión de una narrativa que prioriza la resolución problemas sociales a través de soluciones técnicas (Powles y Nissenbaum, 2018) y asume que la eliminación de las dinámicas de opresión/subordinación entre grupos sociales es un objetivo asumible cuando la lucha contra la discriminación es constante, pero que nunca llegará a un punto culminante de resolución, dado que, intrínsecamente, es una lucha de fuerzas.

La lógica estándar de la ciencia de datos se basa en la precisión y la eficiencia, y tiende a trabajar en el marco de los parámetros de los sistemas existentes, a los que acepta sin cuestionamiento alguno (Green, 2021: 256). Es decir, se considera que las «buenas» decisiones son aquellas que dan respuestas precisas al objetivo dado. Ahora bien, si la atención se limita a la precisión, es casi inevitable que surjan problemáticas relativas a la legitimidad de las decisiones. ¿Es legítimo perfeccionar la precisión de un sistema de vigilancia masiva para que reconozca mejor a las mujeres y a las personas no blancas? El objetivo de que un sistema funcione con más precisión no es necesariamente correcto desde el punto de vista normativo (West, 2020: 3). En contraste, podrían considerarse «buenas» decisiones aquellas que se centraran en las cualidades de los sujetos o en la obtención de resultados más inclusivos, que considerasen solo el conjunto completo de factores relevantes, que

incorporasen principios normativos (por ejemplo, los de necesidad o proporcionalidad), que permitiesen que las personas comprendieran el sistema y pudieran impugnarlo (Barocas, 2019: 33) o que prevean buenas protecciones procesales. Sin embargo, aunque resulta posible introducir protecciones procesales en los sistemas automatizados para justificar su legitimidad, ello implicaría socavar el ahorro de costes que la automatización pretende lograr (Barocas *et al.*, 2019: 43).

3.3. DESCONTEXTUALIZACIÓN

En este apartado pretendemos explicar, en primer lugar, que el hecho de desvincular la discriminación de su dimensión colectiva implica obviar también la historia de minusvaloración que han sufrido las personas perjudicadas, que el tratamiento de datos a través de la IA es un gran método para manejar información ignorando el contexto en el cual esta ha sido extraída (3.3.1). Además, estudiaremos el modo a través del cual la descontextualización del concepto de «discriminación» ha influido en el desarrollo de la «discriminación inversa» (3.3.2).

3.3.1. Tratamiento de la discriminación obviando la dimensión colectiva

Otro aspecto primordial para abordar la discriminación jurídicamente desde la perspectiva antisubordinatoria es el reconocimiento de que el colectivo oprimido debe situarse en el centro de la protección jurídica. El trato desigual o injusto es experimentado por personas individualmente consideradas, pero la razón de este trato es que se les atribuyen ciertas características, rasgos o prejuicios propios de una colectividad (Añón Roig, 2013: 134) independientemente de la heterogeneidad interna del grupo, y que la opresión hacia el grupo se traduce en situaciones de discriminación particulares (Barrère Unzueta, 2008: 60-62). La discriminación tiene una dimensión colectiva y grupal definitoria y no eliminable. No solo eso, sino que, tal y como señala Álvarez del Cuvillo (2022: 191), la prohibición de la discriminación basada en la dimensión colectiva parece ser la única perspectiva que se sostiene desde una interpretación histórica, sociológica, teleológica, sistemática e incluso literal del derecho positivo vigente. En suma, de acuerdo con la perspectiva antisubordinatoria, la discriminación es la manifestación individualizada de sistemas de opresión y dominación. La opresión es una condición de grupos y esta caracterización resulta fundamental (Barrère Unzueta y Morondo Taramundi, 2011: 20).

Sin embargo, el principio de igualdad ante la ley (la cláusula de igual protección) ignora determinados fenómenos estructurales que subyacen a las situaciones injusticia social. Dado que aquel principio está basado en la igualdad de trato individual, no solo deja sin resolver la parte estructural del conflicto, sino que incluso contribuye a perpetuar la injusticia social (Barrère Unzueta, 2008: 58). La disociación entre la normativa antidiscriminatoria y la lucha contra las desigualdades sistémicas que realmente operan en la sociedad constituye una distorsión manifiesta de su finalidad que puede acabar desactivando su operatividad, dado que una visión puramente individualista concibe las causas de discriminación enun-

ciadas en las normas antidiscriminatorias simplemente como clasificaciones abstractas (Álvarez del Cuvillo, 2022: 192).

Este proceso de abstracción invisibiliza las desigualdades realmente existentes en la sociedad, proyectándolas sobre un plano lógico-formal desconectado de la realidad material concreta. Semejante forma de proceder permite concebir la discriminación como una situación anómala, excepcional o patológica, ocultando así situaciones cotidianas que reproducen las relaciones de dominación y subordinación que acaban legitimando el orden de poder desigual (Álvarez del Cuvillo, 2022: 192 y 193). En esta línea, Barrère Unzueta y Morondo Taramundi (2011: 39) señalan que, desde aquel enfoque, la discriminación se percibe como un fenómeno excepcional o accidental que reclama una respuesta puntual del Derecho, cuando en realidad se trata de un fenómeno sistémico. Por su parte, Sebastiani (2021: 756) observa que, en relación con el racismo, si bien el debate hegemónico bascula entre la consideración de la discriminación racial como una consecuencia de prejuicios difundidos y su caracterización como un resultado de predisposiciones patológicas y extremistas, en realidad concibe los episodios racistas como actos «aislados» imputables a individuos concretos y desconectados de dinámicas sociales más profundas. Posiblemente, uso de la IA ha servido para evidenciar el carácter sistémico y difuso de la discriminación y para debilitar la idea de que la responsabilidad radica en un individuo en concreto, dificultando así el planteamiento de respuestas por parte del sistema jurídico, todavía deudor de una perspectiva centrada la responsabilidad individual.

Desde nuestra perspectiva, en el momento en que se pierde de vista el grupo, también se pierde de vista el contexto. La corriente anticlasificatoria desvincula completamente la decisión del contexto, es decir, ignora todas las dinámicas sociales —no solo históricas, sino también actuales— que determinan la previsión jurídica de un atributo protegido. En cuanto a la discriminación algorítmica, los elementos que están implicados en el trato equitativo (por ejemplo, la consideración de la existencia de la discriminación estructural) rara vez se tienen en cuenta en la toma de decisiones concretas, ya que aquellos solo tienen en cuenta causas relativamente próximas a los puntos de decisión (Barocas *et al.*, 2019: 185). Por ejemplo, cuando se toman decisiones con base en el expediente académico, no se tiene en cuenta el hecho de que las diferencias en los resultados a menudo surgen de la inestabilidad en el hogar y otras circunstancias sociales, económicas y de salud (Barocas *et al.,* 2019: 158). El tratamiento de datos se basa, por tanto, en la descontextualización, es decir, traduce conceptos complejos en simples datos descontextualizados. De la misma manera, las clasificaciones que se realizan mediante el análisis masivo de datos pretenden imponer orden en el desorden de la vida humana, pero al hacerlo ocultan la distribución desigual de sus impactos en las comunidades (West, 2020: 4).

La lucha contra los sistemas de opresión y las dinámicas de subordinación que operan en la sociedad exige ir más allá de las simplificaciones y analizar las sutilezas involucradas en las relaciones y las decisiones humanas. La IA hace lo contrario, puesto que el coste de tomar el mayor núnero de decisiones en el mínimo tiempo posible y con el mínimo coste (incremento del valor cuantitativo) es la simplificación de los factores considerados a la hora de tomar la decisión (disminución valor cualitativo). Podríamos decir, siguiendo a

Barocas *et al.* (2019: 25), que estos fatores ponen en relación el aprendizaje automático con la burocracia, que la IA sobresale en la simplificación de los procesos para automatizar las decisiones, a despecho de la pérdida cualitativa que ello conlleva. Indefectiblemente, el traslado de la perspectiva burocrática a todos los ámbitos de adopción de decisión implica indefectiblemente una pérdida de riqueza en estos procesos que afecta más intensamente a colectivos oprimidos, dado que su participación social se está más restringida (Ranchordás y Scarcella, 2022: 418). Además, lo que hace la IA es adoptar los sistemas de dominación/subordinación dominantes (patrones, reglas y normas) y extrapolarlos a todos los ámbitos de toma de decisión, ignorando que contextualizar es una tarea que las personas sí pueden llevar a cabo. Por ejemplo, las personas que viven cerca de una intersección pueden tener más probabilidades de sufrir accidentes porque aquella está mal diseñada y, por tanto, es peligrosa. Esta contingencia raramente podrá ser tenida en cuenta por un sistema algorítmico. El ahorro de costes que podría lograrse a través de la automatización de ciertas decisiones (sobre todo, reemplazando a los trabajadores por *software*) se hace a expensas de privar a las personas de la oportunidad de resaltar aspectos relevantes de la información que no tienen cabida en el proceso automatizado (Barocas *et al.*, 2019: 37). Adicionalmente, cuando el sistema en sí es injusto, las personas encargadas de implementarlo pueden ser una fuente importante de resistencia ante el incumplimiento o la denuncia de irregularidades (Barocas *et al.*, 2019: 217).

3.3.2. Especial mención a la discriminación inversa

El enfoque de la discriminación predominante en el CEDH (Convenio Europeo de Derechos Humanos) y la jurisprudencia del TEDH (Tribunal Europeo de Derechos Humanos) se sitúa claramente en la órbita del principio de igualdad formal. El tratamiento de la discriminación en la normativa antidiscriminatoria de la UE —así como en la jurisprudencia del TJUE (Tribunal de Justicia de la Unión Europea)— es especialmente ambiguo y complejo en lo que respecta a la integración de las perspectivas individual y colectiva: aunque reconoce expresamente la legitimidad de medidas como las acciones positivas, que solo pueden entenderse desde una perspectiva antisubordinatoria, tiende a identificar las causas de discriminación con las clasificaciones contrarias al principio general de igualdad de trato entendido en términos abstractos —una comprensión que, por tanto, permite la tutela simétrica—. De hecho, la jurisprudencia del TJUE admite abiertamente la posibilidad de discriminación inversa, dado que ha tomado en consideración a menudo ha atendido las reclamaciones basadas en la supuesta discriminación de miembros de grupos dominantes (*vid.,* por ejemplo, las sentencias *Eckhard Kalanke c. Freie Hansestadt Bremen*, de 17 de octubre de 1995 (C-450/93), y *WA c. Instituto Nacional de la Seguridad Social* (INSS), de 12 de diciembre de 2019 (C-450/18)) (Álvarez del Cuvillo, 2022: 195).

La admisión de la discriminación inversa se puede entendida como una consecuencia de la omisión del contexto en el que se produce la discriminación. Generalmente, la discriminación es caracterizada como una figura híbrida que abarca tanto una dimensión individual como una dimensión colectiva. A pesar de que existen algunos puntos de con-

vergencia entre ambas dimensiones, en realidad hacen referencia a paradigmas contradictorios y mutuamente excluyentes, lo que no impide que se apliquen en la práctica de forma combinada al precio de incurrir numerosas incoherencias y contradicciones lógicas. Una de estas contradicciones es el concepto de «discriminación inversa» que se aplica a aquellos tratos que generar un perjuicio a personas que pertenecen a grupos sociales dominantes o mayoritarios precisamente por su adscripción a estos colectivos. Se trata de una apropiación de la categoría jurídica de la discriminación llevada a cabo por los grupos dominantes que acaba diluyendo su potencial transformador y legitimando así el orden de poder desigual. En estos supuestos, la lucha contra la discriminación no solo deja de cumplir su función emancipadora, sino que, además, se convierte en un instrumento de dominación (Álvarez del Cuvillo, 2022: 190, 191 y 196). Barocas y Selbst (2016: 723-728) explican que en varios litigios de Estados Unidos de América los órganos de adjudicación han redefinido la distinción entre categorizaciones buenas y perjudiciales de raza reconduciéndola al principio formalista de anticlasificación; de este modo, los tribunales suprimen la función originaria del derecho antidiscriminación —proteger a miembros de grupos históricamente desaventajados— y facilitan que los hombres blancos puedan denunciar que han sufrido discriminación. Al decir de los autores, el fallo recaído en el caso *Ricci vs. DeStefano* 557 U.S. 557 (2009) constituye un claro ejemplo de esta renovada línea jurisprudencial.

La perspectiva antisubordinatoria está claramente confrontada con el individualismo imperante en la sociedad y sostiene que para cambiar los patrones históricos de subordinación es preciso cambiar dinámicas de privilegio que se perciben como «justas». Por ejemplo, que la acción positiva genera animadversión en grupos históricamente privilegiados porque consideran que no es «justa» o que están pagando ellos por opresiones históricas que ahora ya no existen. Si no hay un cambio en la percepción de lo que se considera justo (por ejemplo, la confrontación entre la simple meritocracia y la toma de decisiones teniendo en cuenta el contexto de la persona para conseguir esos méritos —que conlleva un análisis mucho más complejo—), las personas individualmente perjudicadas por las políticas antisubordinatorias grupales (por ejemplo, aquellas que no pueden acceder a un trabajo que consideran que deberían obtener porque es lo más justo desde la perspectiva de los grupos sociales privilegiados que históricamente han establecido los criterios a partir de los cuales se toman las decisiones), van a oponer una fuerte resistencia a la adopción e implementación de estas políticas, resistencia cuya razón de ser radica en su capacidad de abstraerse y desvincularse del contexto social opresivo cuando las decisiones les afectan individualmente.

La conducta discriminatoria está conformada por la confluencia de dos elementos constitutivos: por un lado, la pertenencia a uno o varios grupos o categorías sociales, y, por otro, el perjuicio individual y social provocado por el tratamiento derivado de esta adscripción, que lesiona la dignidad humana porque tiende a situar a determinados grupos en una posición socio-jurídica de inferioridad. Así pues, solo serían discriminatorios los tratamientos desiguales que estuvieran funcional o sistemáticamente orientados a la marginación de determinados grupos sociales y las personas que los integran. Para que una conducta sea considerada discriminatoria, se exige que pueda ser contemplada como la manifestación de

una pauta sistemática de degradación de los integrantes de un grupo que, de generalizarse o reproducirse, contribuye a situarlos en una posición de inferioridad social (Álvarez del Cuvillo, 2022: 190, 191 y 196). Este modelo teórico parte de la base de que todo acto de discriminación debe causar un perjuicio a determinados grupos humanos, aunque el perjuicio sea solo potencial (Álvarez del Cuvillo, 2022: 208). Desde el momento en que se admite que una norma puede tener un trasfondo discriminatorio, también se admite que no se trata de proteger solo a las personas discriminadas, sino también a los grupos o colectivos que sufren un perjuicio por sus condiciones personales o sociales (Rodríguez- Piñero y Bravo-Ferrer, 2022: 15).

3.4. MONOCULTURA ALGORÍTMICA

La búsqueda de la equidad algorítmica formal, entendida como la garantía de un trato igual y no arbitrario por parte de los sistemas algorítmicos, es un compromiso con las personas individualmente consideradas y no con los grupos sociales a los cuales pertenecen, ya que para adoptar un compromiso de equidad en relación con ciertos colectivos sociales subordinados debe existir un trato diferenciado en favor de estos. Green (2022: 3) plantea que es imposible que un algoritmo satisfaga todas las definiciones matemáticas en la adopción de una decisión equitativa.

En lugar de plantear la necesidad de dispensar un trato favorable a los grupos históricamente (y actualmente) subordinados, el camino que se está siguiendo es el de uniformizar las decisiones, dado que los sistemas de aprendizaje automático que se están generalizando son los de la cultura mayoritaria (basada en sistemas de opresión y subordinación), sistemas que generan altas tasas de error que perjudican a los grupos minoritarios (Barocas *et al.*, 2019: 12). Kleinberg y Raghavan (2021: 1) llaman a esta situación «monocultura algorítmica», noción que hace referencia al hecho de que las elecciones y preferencias devienen homogéneas cuando se ven implicados sistemas algorítmicos, ya que, si muchos sistemas de aprendizaje automático utilizan los mismos datos de entrenamiento y la misma variable objetivo, tenderán a generar las mismas clasificaciones, incluso si los algoritmos de aprendizaje son muy distintos. Por ejemplo, un pequeño número de empresas de medios sociales determinan en conjunto qué tipos de discurso pueden formar parte del discurso principal en línea y qué comunidades pueden movilizarse en línea (Barocas *et al.*, 2019: 215). Por otro lado, cuando se utiliza el mismo algoritmo o diferentes algoritmos basados en los mismos datos en múltiples contextos, una persona puede ser excluida arbitrariamente de una amplia gama de oportunidades (Creel y Hellman, 2021: 35 y 36).

Un ejemplo de interés en relación con el impacto de los sistemas algorítmicos en culturas minoritarias son los sistemas de reconocimiento de voz, que son menos precisos cuando se les habla en dialectos locales, circunstancia que motiva que cada vez sea más común ajustar la pronunciación y neutralizar el acento y la entonación, es decir, adoptar un comportamiento autocorrectivo para asegurar que el *software* de reconocimiento de voz recoja las palabras correspondientes integrando los sesgos del aprendizaje automático a la propia sociedad (Pasquinelli, 2022: 27).

Hay que tener presente que el hecho de haber ostentado históricamente (y actualmente) el poder de crear normas que rijan la sociedad implica que se hayan creado unos estándares con base en los cuales se medirá la igualdad; el proceso para garantizar la igualdad de derechos y oportunidades puede ser, por ello, confundido con un proceso para asimilar los usos y particularidades de los grupos sociales subordinados a los usos y particularidades de los grupos privilegiados. Mackinnon (1991:82) planteó que la vigencia del estándar de acuerdo con el cual todas las personas son iguales provoca que las mujeres se midan según la correspondencia con los hombres: por consiguiente, la neutralidad es simplemente el estándar masculino. Para apreciar la existencia de discriminación, es necesario que existan diferentes personas o grupos de personas que son tratados de forma diferente, y para probar su existencia se ha requerido un término comparativo, es decir, un referente comparativo que se erige como modelo de un trato privilegiado frente a otro. Esta concepción ha suscitado la crítica que apunta al «asimilacionismo», entendido como la absorción y convalidación implícita del modelo con el que se compara (Barrère, 2001: 151). Holtmatt (2010: 203) plantea que el hecho de no integrar el trato desigual a las mujeres cuando se las compara con los hombres —y asimilar, por tanto, las necesidades y preocupaciones de las mujeres a la de los hombres— deja intactos los patrones masculinos ya existentes. En la misma dirección apunta Green (2022:10) cuanto sostiene que, si ciertos colectivos sociales se enfrentan a oportunidades de desarrollo muy diferentes (que generalmente dependen de atributos que se distribuyen de manera desigual entre los grupos debido a la opresión, agravando las desventajas existentes), es imposible crear sistemas competitivos justos, de manera que el objetivo no debería limitarse a ayudar a algunas personas que forman parte de grupos sociales subordinados a que reciban decisiones favorables a través de un trato especial, aceptando como dada la estructura de oportunidades, sino cambiar la propia estructura de oportunidades. En este sentido, y para considerar métodos que favorezcan la aplicación de medidas contra la opresión en los sistemas algorítmicos hay que tener en cuenta que la pretensión de los grupos sociales subordinados es obtener el mismo respeto social, independientemente de sus capacidades y atributos, y no ser situados en los peldaños inferiores de una escala jerárquica. No obstante, esta pretensión no debe ser confundida con la voluntad de absorber y convalidar implícitamente las dinámicas, los roles y los estándares atribuidos al grupo privilegiado, que es el que promueve —en el caso de la IA, de manera exacerbada— unos estándares y una cultura concretos. Es decir, el objetivo es conseguir un respeto equivalente entre grupos sociales, no la asimilación de los roles, las actitudes y la cultura de los grupos sociales privilegiados que tradicionalmente han operado como estándares.

4. MEDIDAS ANTISUBORDINATORIAS A APLICAR

Siguiendo la perspectiva según la cual la lucha contra la discriminación es constante que nunca llegará a un punto culminante de resolución porque es una lucha de fuerzas intrínseca, proporcionar soluciones implica convencer a aquellas personas que forman parte

de los grupos sociales privilegiados de que las prerrogativas que entienden como «justas» y «no arbitrarias» únicamente lo son a la luz de sus estándares, así como de la necesidad de llegar a acuerdos sobre el modo de desarrollar formas de razonamiento para juzgar cuando y qué cantidad de discriminación es tolerable (Barocas y Selbst, 2016: 728). Hay que tener en cuenta, especialmente en un campo como el de la IA, en el que la inmediatez es tan relevante, que las medidas antisubordinatorias están concebidas para que tengan una incidencia a largo plazo, circunstancia que puede propiciar que se las considere poco atrayentes para abordar los problemas de discriminación que se plantean en sistemas algorítmicos concretos con base en los cuales se pretende obtener una rentabilidad coste-beneficio.

En relación con los datos, para paliar los problemas de equidad que presentan los resultados que generan los sistemas algorítmicos se ha planteado la posibilidad de recopilar más datos y realizar inversiones para mejorar la tecnología de clasificación a fin de de mitigar potencialmente la disparidad en la tasa de error (Barocas *et al.*, 2019: 98), o divulgar masivamente datos sociales que sustenten el desarrollo de la IA teniendo en cuenta que ya es tarde para hacer una reevaluación radical del modo en que las grandes empresas tecnológicas controlan las enormes cantidades de datos que se utilizan para desarrollar la IA (Powles y Nissenbaum, 2018). Estas medidas se plantean después de aceptar que los algoritmos de aprendizaje automático utilizan y han utilizado bases de datos masivos, que normalmente han obtenido los datos que revelan información sensible e íntima sobre las personas de manera ilícita o con base en un consentimiento no significativo (Green, 2021: 249). Por un lado, podría interpretarse que las normativas de protección de datos han intentado regular el uso exacerbado de datos para crear información útil con finalidades comerciales o de eficiencia en los procesos de decisión. Por otro, también se puede interpretar que la laxitud de las normativas de protección de datos ha permitido el desarrollo sin control de la IA. Al respecto, es interesante la propuesta de Jo y Gebru (2020), consistente en extraer lecciones de la archivística y la bibliotecología para documentar conjuntos de datos de aprendizaje automático prestando atención a cuestiones de consentimiento, inclusión, poder, transparencia, ética y privacidad.

En relación con la legitimidad, antes de utilizar la toma de decisiones mediante sistemas algorítmicos sería interesante plantear cuál será el impacto que tendrá en la persistencia y la magnitud de los sistemas sociales de opresión y subordinación. En este sentido, se ha presentado la idea de reconocer a determinados colectivos sociales —por ejemplo, comunidades geográficas— el derecho a consentir o rechazar colectivamente la adopción de herramientas tecnológicas (Barocas *et al.,* 2019: 219). Se plantea aquí un escenario en el que tanto las personas que deciden de qué manera se van a tomar las decisiones como las personas afectadas por estas decisiones disponen de margen para actuar. En muchos casos, si no en la mayoría, no va a ser así, de manera que no es descabellado plantear la prohibición normativa del uso de sistemas de IA en contextos sociales si perpetúan y agraven los sistemas de opresión y desigualdad, o que se someta a los sistemas de IA que se utilicen en el sector público —al menos— a un escrutinio exhaustivo y de prueba en relación con el impacto que pueda tener su uso en los colectivos desaventajados de la sociedad.

En relación con los estereotipos, Pou (2016: 175) plantea que para examinar una acción o norma presuntamente discriminatoria se incluya el razonamiento sobre si esta crea, perpetúa o agrava alguno de los tratos que mantienen cierto tipo distintivo de desventaja. En la misma dirección, Ghidoni y Morondo Taramundi (2022: 41 y 42) ofrecen ejemplos de medidas orientadas a reducir el impacto de los estereotipos en los resultados de los procesos decisorios (extrapolables al campo de la IA), a diversificar la composición social, de género, racial, de edad o clase de las personas que se ocupan del análisis de datos, a exponer a las personas que tienen que tomar decisiones a contra-narrativas e imágenes positivas de los grupos estereotipados. Asimismo, dan cuenta de diversas estrategias de sensibilización, entre ellas las técnicas que promueven la «autocorrección» en la toma de decisiones, que podría incluirse en los sistemas algorítmicos a través de procesos de revisión por parte de personas para discernir entre los casos en los que el sistema algorítmico ha encontrado patrones que representan conocimientos o los que representan estereotipos (nos remitimos a la ejemplificación que hemos presentado en el apartado 3.1). Por otra parte, las autoras sostienen que, si bien el Derecho está compuesto por normas abstractas (y generales) y funciona a través de categorías, esto no significa que no sea posible discutir sobre las formas de generalización del Derecho, pues este es un ámbito idóneo para introducir la discusión sobre los estereotipos (*ibid.*: 43). Aunque en el ámbito del Derecho —un campo más controlado en el que existen procesos de decisión pautados— la discusión puede ser fructífera, parece que esta posibilidad es más complicada en el ámbito de la IA, terreno en el que la generalización se produce de manera menos calculada y no hay tantos controles en los que puedan introducirse criterios orientados a luchar contra los estereotipos. Por tanto, es más útil sensibilizar directamente las personas que van a trabajar en el campo de la IA (por ejemplo, cuando cursan sus estudios universitarios).

Es necesario reevaluar la posición de los mecanismos de análisis masivo de datos en la sociedad evitando de la presunción de acuerdo con la cual el aprendizaje automático mejora los procesos decisorios en cualquier caso (Green, 2021: 257). También genera suspicacias el hecho de que se plantee como una alternativa para mejorar los procesos decisorios cuando, en el ámbito de las relaciones y las estructuras sociales, parece que lo que hace es perpetuar el *statu quo* opresivo a gran escala.

5. CONCLUSIONES

Actualmente predomina la perspectiva que concibe el principio de igualdad de trato como indiferenciación o trato razonable no arbitrario y esta preponderancia tiene consecuencias para el abordaje de la discriminación algorítmica. Hemos analizado cuatro elementos clave de la perspectiva antisubordinatoria que tienen efectos en las vías a través de las que se pretende mitigar la discriminación que producen los sistemas algorítmicos. En primer lugar, hemos mostrado que el hecho de no tener en cuenta los estereotipos que operan en la sociedad —que impregnan también el ámbito de la IA— no solo tiene como consecuencia que los métodos creados para hacer frente a la discriminación algorítmica

puedan ser inefectivos, sino que también que estos puedan contribuir a perpetuar la discriminación. En segundo lugar, hemos puesto en evidencia que la neutralidad que se atribuye tanto al Derecho como a la IA solo es aparente y tiende a promover la conformidad con los valores sociales y políticos dominantes. En tercer lugar, hemos analizado las implicaciones asociadas a la comprensión del Derecho antidiscriminatorio desde la perspectiva anticlasificatoria, entre ellas la abstracción de la realidad y la omisión del contexto histórico-social en el que se producen las discriminaciones. En el análisis masivo de datos también se produce esta descontextualización, pues aquellos se utilizan como información sin tener en cuenta de dónde proviene. El hecho de que en estas circunstancias no se contextualice adecuadamente puede contribuir a robustecer la idea de que la dimensión colectiva de la discriminación es poco relevante, y facilitar que se aprecien casos de discriminación en los cuales la persona perjudicada no pertenece a un colectivo socialmente subordinado. En cuarto lugar, hemos explicado que la toma de decisiones mediante sistemas algorítmicos (que, al difundirse, interiorizan estándares de la cultura mayoritaria) supone un incremento en la uniformidad de las elecciones y preferencias que tiene una incidencia directa negativa en la consideración y prevalencia de los usos y culturas de grupos subordinados.

Los elementos estudiados generan preocupaciones desde la perspectiva antisubordinatoria tanto en el ámbito del Derecho como en el de la IA, dado que son dos esferas en las que se concentra el poder y que han estado y están dominadas por colectivos sociales privilegiados. La perspectiva social y cultural de los hombres ha construido la totalidad de un ordenamiento jurídico que se pretende neutral, del mismo modo que la perspectiva social y cultural de los ha construido la totalidad de una «inteligencia» (énfasis en el entrecomillado) artificial que pretende aparentar neutralidad. Es importante tener en cuenta que los sistemas algorítmicos magnifican la discriminación que opera en la sociedad, ya que tienen un impacto potencial sobre muchas personas, y que las soluciones que se están proponiendo para abordar la discriminación que generan se limitan a los mismos mecanismos tradicionales que se han utilizado hasta ahora para hacer frente a la discriminación, mayoritariamente inspirados en el principio de igualdas formal. En la sección 4 hemos presentado algunas propuestas orientadas a evitar la perpetuación de las dinámicas de opresión/subordinación causadas por los sistemas algorítmicos, entre ellas la formación contra los estereotipos de las personas que se dedican al análisis de datos o la prohibición normativa del uso de sistemas de IA en contextos sociales en aquellos casos en los que perpetúen y agraven (o puedan perpetuar o agravar) la opresión y la desigualdad.

En cualquier caso, es necesario preguntarse si, en términos normativos, el objetivo es conseguir que los sistemas de IA se limiten a no incrementar la discriminación existente en la sociedad o si lo que se pretende es que estos sistemas también participen en su corrección. La respuesta a esta cuestión conlleva la adopción de diferentes medidas y diferentes perspectivas de análisis.

Desde nuestra perspectiva, la utilización de los sistemas de IA contribuye a poner de manifiesto que la discriminación es un problema complejo, acaso más complejo de lo que muestra su tratamiento jurídico por ello, están cobrando relevancia planteamientos hasta ahora más minoritarios. Si bien es cierto que ante una sofisticación tecnológica debe

adoptarse una sofisticación normativa y de análisis de los conflictos (por ejemplo, la anticlasificación está basada en un análisis mucho más simplista de las relaciones de poder y las dinámicas de opresión/subordinación que operan en la sociedad que la antisubordinación). Esta sofisticación no se ha conseguido aún en relación con el Derecho ¿Es posible que se consiga en el campo de la IA? Las medidas antisubordinatorias son aquellas que más lejos han llegado a la hora de conseguir una para los grupos sociales subordinados una renuncia de privilegios por parte de los grupos sociales favorecidos. Las medidas antisubordinatorias difícilmente pueden traducirse en conceptos técnicos y raramente pueden traducirse en resultados concretos a corto plazo que justifiquen la introducción (o no) en el mercado de un bien o servicio de la IA (contrariamente a las medidas anticlasificatorias: por ejemplo, conseguir que el sistema algorítmico ignore ciertos atributos protegidos). A pesar de que pueden revelarse como una buena estrategia para luchar contra la opresión, es necesario que buena parte de la población esté convencida de que su implementación es necesaria, efectiva y deseable.

BIBLIOGRAFÍA

ÁLVAREZ DEL CUVILLO, Antonio (2022): «El problema de la discriminación inversa: ¿Es posible discriminar a quienes pertene cen a los grupos sociales dominantes?», *Trabajo, Persona, Derecho, Mercado,* 5, 187-209.

AÑÓN ROIG, María José (2013): «Principio antidiscriminatorio y determinación de la desventaja», *Isonomía: Revista de Teoría y Filosofía del Derecho,* 39, 127-157.

— (2021): «Transformaciones en el derecho antidiscriminatorio: avances frente a la subordinación», *Revista Electrónica del Instituto de Investigaciones Jurídicas y Sociales Ambrosio Lucas Gioja*, 26, 29-54.

— (2022): «Desigualdades algorítmicas: conductas de alto riesgo para los derechos humanos», *Derechos y libertades,* 47, 17-49.

BARRÈRE UNZUETA, María Ángeles (2001): «Problemas del derecho antidiscriminatorio: subordinación versus discriminación y acción positiva versus igualdad de oportunidades», *Revista Vasca de Administración Pública. Herri-Arduralaritzako Euskal Aldizkaria*, 60, 145-166.

— (2008): «Iusfeminismo y derecho antidiscriminatorio: hacia la igualdad por la discriminación», en R. M. Mestre i Mestre (coord.), *Mujeres, derechos y ciudadanías,* Valencia: Tirant Lo Blanch, 45-71.

— (2018): «Filosofías del Derecho Positivo ¿Qué Derecho y qué discriminación? Una visión contrahegemónica del Derecho Antidiscriminatorio», *Anuario de Filosofía del Derecho,* XXXIV, 11-42.

BARRÈRE UNZUETA, María Ángeles y Dolores MORONDO TARAMUNDI (2011): «Subordinación y discriminación interseccional: elementos para una teoría del derecho antidiscriminatorio», *Anales de la Cátedra Francisco Suárez,* 45, 15-42.

BAROCAS, Solon *et al.* (2019): «Fairness and Machine Learning: Limitations and Opportunities» [en línea] <https://fairmlbook.org/>. [Consulta: 12/02/2024.]

BAROCAS, Solon y Anderw D. SELBST (2016): «Big Data's Disparate Impact», *California Law Review*, 104, 671-732.

BIGMAN, Yochanan E. *et al.,* «Algorithmic Discrimination Causes Less Moral Outrage Than Human Discrimination», *Journal of Experimental Psychology: General*, 152(1), 4-27.

BIRHANE, Abeba *et al. (2022):* «The values encoded in machine learning research», *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 173-184.

COOK, Rebecca J. y Simone CUSACK (2010): *Estereotipos de género. Perspectivas Legales Transnacionales*, Bogota: Profamilia.
CRAWFORD, Kate (2021): *Atlas of AI*, New Haven and London: Yale University Press.
CREEL, Kathleen y Deborah HELLMAN (2022): «The algorithmic leviathan: Arbitrariness, fairness, and opportunity in algorithmic decision-making systems», *Canadian Journal of Philosophy*, 52 (1), 26-43.
FARKAS, Lilla (2014): «Collective actions under European anti-discrimination law», *European Anti-Discrimination Law Review*, 19, 25-40.
GONEN, Hila y Yoav GOLDBERG (2019): «Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But do not Remove Them» [en línea] <https://arxiv.org/pdf/1903.03862.pdf>. [Consulta: 20/02/2024.]
GREEN, Ben (2021): «Data Science as Political Action: Grounding Data Science in a Politics of Justice», *Journal of Social Computing*, 2(3), 249-265.
— (2022): «Escaping the Impossibility of Fairness: From Formal to Substantive Algorithmic Fairness», *Philosophy & Technology,* 35, 1-32.
GHIDONI, Elena y Dolores MORONDO TARAMUNDI (2022): «El papel de los estereotipos en las formas de la desigualdad compleja: algunos apuntes desde la teoría feminista del derecho antidiscriminatorio», *Discusiones,* 28, 37-70.
HOLTMAAT, Rikki (2010): «Equal Treatment to Equal Right», en E. Bodelón y D. Heim (coords.), *Law, Gender and Equality*, Barcelona: Universitat Autònoma de Barcelona, 209-228.
JO, Eun Seo y Timnit GEBRU (2020): «Lessons from archives: Strategies for collecting sociocultural data in machine learning», *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency,* 306-316.
KLEINBERG, Jon y Manish RAGHAVAN (2021): «Algorithmic monoculture and social welfare», *Proceedings of the National Academy of Sciences*, 118(22), 1-7.
LOUSADA AROCHENA, José Fernando (2022): «Evolución de la igualdad desde la constitución de 1978: del patriarcado fuerte hacia la igualdad de género», *IQUAL. Revista de género e igualdad,* 5, 1-27.
MACKINNON, Catherine (1991): «Difference and Dominance: on sex discrimination», en K. Bartlett y R. Kennedy (eds.), *Feminist legal theory: readings in law and gender*, Boulder: Westview Press, 81-94.
MORONDO TARAMUNDI, Dolores (2023): «Los estereotipos como mecanismos de desigualdad y alienación: un análisis desde el derecho antidiscriminatorio», *Oñati Socio-Legal Series,* 13(3), 710-729.
PASQUINELLI, Matteo (2022): «Cómo una màquina aprende y falla. Una gramàtica del error para la Inteligencia Artificial», *Revista Hipertextos,* 10(17), 13-29.
POU, Francisca (2016): «Estereotipos, daño dignitario y patrones sistémicos: la discriminación por edad y género en el mercado laboral», *Discusiones*, 16(1), 147-188.
POWLES, Julia y Helen NISSENBAUM (2018): «The seductive diversion of 'solving' bias in artificial intelligence» [en línea] <https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53>. [Consulta: 12/02/2024.]
RANCHORDÁS, Sofia y Luisa SCARCELLA (2022): «Automated Government for Vulnerable Citizens: Intermediating Rights», *William & Mary Bill of Rights Journal*, 30(2), 373-418.
RIVAS VALLEJO, Pilar (2020): *La aplicación de la Inteligencia Artificial del trabajo y su impacto discriminatorio*, Navarra: Aranzadi.
ROBERTS, Jessica L. (2015): «Protecting Privacy to prevent discrimination», *William and Mary Law Review,* 56(6), 2123-2124.
RODRÍGUEZ-PIÑERO y BRAVO-FERRER, Miguel (2022): «Los contornos de la discriminación», T*emas Laborales,* 162, 11-18.
SCHREURS, Wim *et al.* (2008): «Cogitas, Ergo Sum. The role of data protection Law and non-discrimination law in group profiling in the private sector», en M. Hidelbrandt y S. Gutwirth (eds.)*, Profiling the European citizen: Cross-Disciplinary Perspectives*, Berlin: Springer Dordrecht, 241-270.
SEBASTIANI, Luca (2021): «Investigando los límites de la lucha legal contra el racismo: El marco español de antidiscriminación por origen racial o étnico», *Oñati Socio-legal series*, 11(3), 753-786.

SORIANO ARNANZ, Alba (2021): «Decisiones automatizadas y discriminación: aproximación y propuestas generales», *Revista General de Derecho Administrativo*, 56, 1-45.
WACHTER, Sandra (2021): «How Fair AI Can Make Us Richer», *European Data Protection Law Review*, 7(3), 367-372.
WEST, Sarah Myers (2020): «Redistribution and Rekognition: A Feminist Critique of Algorithmic Fairness», *Catalyst: Feminism, Theory, Technoscience*, 6(2), 1-24.

Fecha de recepción: 3 de julio de 2024.
Fecha de aceptación: 20 de octubre de 2024.