

## PROBAR LO INVISIBLE: LOS DESAFÍOS PROBATORIOS DE LA DISCRIMINACIÓN ALGORÍTMICA\*

Proving the Invisible: Evidentiary Challenges in Algorithmic  
Discrimination

CODDOU-MC MANUS, ALBERTO \*\*

*Pontificia Universidad Católica de Chile*

FISCHER, EMMA \*\*\*

*Universidad Diego Portales*

### Resumen

Este artículo analiza los desafíos probatorios de las víctimas de discriminación algorítmica, considerando principalmente el caso del acceso al trabajo. Examina cinco barreras clave, revisa normas comparadas y destaca la necesidad de marcos normativos que alivianen o redistribuyan la carga probatoria.

### Palabras clave

Discriminación algorítmica; prueba; reclutamiento y selección.

### Abstract

This article examines the evidentiary challenges faced by victims of algorithmic discrimination, focusing primarily on access to employment. It identifies five key barriers, reviews comparative regulations, and highlights the need for legal frameworks that ease or shift the burden of proof.

### Key words

Algorithmic discrimination; evidence; recruitment and selection.

## 1. Introducción

El presente trabajo comenzará con un relato ficticio pero verosímil en las actuales condiciones del desarrollo de nuevas tecnologías digitales. Una joven abogada pretende comenzar su carrera en alguno de los estudios de abogados más famosos del país. Una amiga le comenta acerca de una plataforma especializada en la que ella puede subir su currículum y postular a las distintas posiciones laborales disponibles en los estudios de abogados. Eventualmente, algunos estudios pueden contactarla directamente de acuerdo con los perfiles laborales que están buscando. Además de su currículum, ella debe realizar un *test* de personalidad que está disponible en la plataforma de manera gratuita y tiene la posibilidad de realizar otros exámenes que permitan acreditar el dominio de ciertos conocimientos (por ejemplo, su nivel de inglés) o de otras herramientas (por ejemplo, su destreza en el uso de

\* Este trabajo fue financiado por ANID/FONDECYT/Regular 1230895; ANID/FONDECYT/Iniciación 11220370. Los autores declaran sus contribuciones al presente artículo de investigación, utilizando la taxonomía CRedit: Conceptualización: Alberto Coddou (80%), Emma Fischer (20%); Investigación: Alberto Coddou (70%), Emma Fischer (30%); Redacción - borrador original: Alberto Coddou (60%), Emma Fischer (40%); Redacción - revisión y edición: FAlberto Coddou (60%), Emma Fischer (40%).

\*\* College / Escuela de Gobierno, Pontificia Universidad Católica de Chile, Santiago, Chile. Correo electrónico: acoddou@uc.cl; ORCID: <https://orcid.org/0000-0003-2041-2304>.

\*\*\* Universidad Diego Portales, Santiago, Chile. Correo electrónico: emma.fischer.g@gmail.com; ORCID: <https://orcid.org/0009-0001-3837-0643>.

planillas de cálculo), o incluso subir videos de presentación que reemplazan las cartas de presentación. Después de completar todas estas etapas, ella decide postular a tres estudios, además de un cuarto que le envía un mensaje directo invitándola a postular. En dos de estos procesos ella pasa a las etapas siguientes: en uno de ellos, la joven abogada debe someterse a una entrevista videogramada con preguntas predeterminadas que le realiza un sistema equipado con inteligencia artificial; en el otro, ella debe “jugar”, de manera asincrónica, un videojuego de roles que la posicionan en situaciones incómodas o difíciles. Al someterse a estos procesos, ella manifiesta cierta incomodidad, pero decide esperar el resultado de las postulaciones. Confía en su breve pero sólida trayectoria académica y profesional, además de sentirse aliviada de no haber sido entrevistada por abogados que en el pasado le habían hecho comentarios inapropiados. Finalmente, y después de esperar algunos días, recibe la triste noticia de que no ha sido seleccionada en ninguno de los procesos. La misma amiga que le había comentado acerca de esta plataforma de reclutamiento y selección de jóvenes talentos en la industria jurídica le cuenta que en esos procesos han sido seleccionados algunos de sus ex compañeros de carrera que no eran particularmente estudiados. Desde ese mismo día, empieza a sospechar que estos procesos la terminaron discriminando por su género, pero no tiene mucha claridad acerca de qué pasos debería seguir para aclarar sus sospechas. La primera medida que adopta es estudiar los términos y condiciones que irreflexivamente aceptó cuando descargó la aplicación de la plataforma de reclutamiento, pero no encuentra nada que le llame la atención pues se encontraban todas las cláusulas tradicionales sobre protección de datos personales y garantías de confidencialidad. Aun no sabe cómo seguir con su investigación, pero un nuevo antecedente le genera cierta sospecha. Navegando en los sitios web de los estudios de abogados a los que postuló, se entera que los perfiles de los jóvenes abogados son significativamente parecidos: la mayoría de ellos, hombres, provenientes de las mismas universidades y colegios, además de compartir hobbies y otros intereses. A pesar de ello, no tiene certezas acerca de cómo seguir con su investigación, y decide dejar de insistir con su tesis que ya le ha costado algunas burlas de amigos y amigas que la consideran exagerada.

Este relato, si bien adornado acomodaticiamente para el desarrollo de este trabajo, da cuenta de un fenómeno masivo y extendido en diversas industrias: el reclutamiento y selección de personas a través de procesos automatizados. Este es un fenómeno que ilustra uno de carácter más general, y que tiene lugar en diversos ámbitos, tanto públicos como privados. En efecto, diversas decisiones fundamentales que impactan la vida de las personas son adoptadas total o parcialmente por tecnologías digitales. Estas tecnologías utilizan procesos automatizados que se basan en la recolección, gestión y análisis de datos. Estos datos, a su vez, transforman la información relevante en lenguaje abstracto, o código, que estos sistemas pueden traducir e interpretar para tomar una decisión. A través de algoritmos – un conjunto de instrucciones o pasos predefinidos para procesar datos y producir un *output* – diferentes ámbitos están entregados a procesos que aplican tecnologías digitales que no podrían desarrollarse “a mano” o con dispositivos analógicos. En el ámbito público, existen múltiples programas y beneficios sociales que hoy en día son gestionados y provistos a través de procesos automatizados de decisión con poca o casi nula participación humana. Beneficios que antes eran provistos por profesionales especialmente entrenados para ello, y por burocracias que debían seguir reglas que suponían un importante costo de transacción, hoy en día son otorgados por algoritmos que se utilizan para la adopción de decisiones públicas<sup>1</sup>. En el ámbito privado, estos procesos han impactado significativamente en el campo laboral, como lo comprueba la gran cantidad de actividades de instrucción, control y evaluación que hoy en día son realizadas por tecnologías digitales que han venido a reemplazar las relaciones humanas que usualmente configuraban la actividad laboral<sup>2</sup>.

---

<sup>1</sup> ALSTON (2018), pp. 7-8.

<sup>2</sup> OIT (2021), p. 4.

El uso de algoritmos en los más diversos ámbitos ha tenido efectos disruptivos y transformativos en la vida de las personas, generando efectos positivos y negativos<sup>3</sup>. Por un lado, la cantidad de información que puede ser gestionada por tecnologías digitales ha aumentado significativamente la eficiencia y disminuido los costos de transacción de variadas actividades, como la selección de personas, la utilización de bienes de capital o el control de reglas burocráticas. Por otro, sin embargo, se han suscitado diferentes problemas por el impacto que se puede generar en las fuentes de trabajo y en las condiciones laborales (automatización del trabajo), o por la falta de responsabilidad ante los posibles daños que se pueden derivar de decisiones algorítmicas.

Entre los riesgos o impactos negativos de las tecnologías digitales que han recibido mayor atención recientemente, la discriminación algorítmica ha ido adquiriendo un lugar fundamental. La ilusión de que las decisiones algorítmicas podrían cumplir con su promesa de evitar la arbitrariedad y los prejuicios que usualmente abundan en las decisiones adoptadas por seres humanos se ha cuestionado por un robusto cuerpo de literatura académica<sup>4</sup>. De acuerdo con un análisis de los riesgos globales, realizado por el Foro Económico Mundial, la “desigualdad digital” está en el top 10<sup>5</sup>. En el último tiempo, las regulaciones relativas a la protección de datos o las políticas de IA de diversas jurisdicciones destacan la necesidad de proteger la “igualdad digital”<sup>6</sup>, prevenir los “sesgos algorítmicos”<sup>7</sup> o luchar contra todas las formas de “discriminación algorítmica”<sup>8</sup>. Según Fjeld y otros, la gran mayoría de los documentos oficiales que buscan desarrollar estándares a ser aplicados en los procesos automatizados de decisión que implican técnicas de IA mencionan el principio de igualdad y no discriminación entre sus directrices principales<sup>9</sup>. Estas menciones, sin embargo, no distinguen las formas particulares de discriminación que se producen en el marco de procesos automatizados de decisión en ámbitos públicos o privados. Menos aun, estas menciones dedican algún espacio a los complejos desafíos probatorios que deben sortear las víctimas de discriminación algorítmica.

En este escenario, y considerando la insuficiencia del derecho de protección de datos para prevenir y sancionar todas las afectaciones a derechos fundamentales derivadas del uso de nuevas tecnologías digitales<sup>10</sup>, el objetivo principal del artículo consiste en analizar las diferentes formas de discriminación algorítmica y mapear los principales desafíos probatorios que surgen para las víctimas. Para ello, y a lo largo del artículo, volveremos al relato inicial para mostrar al lector las dificultades prácticas que debería sortear una víctima de discriminación algorítmica. En el desarrollo de nuestro trabajo, nos vamos a referir a la normativa y jurisprudencia judicial y administrativa pertinente en regímenes comparados, teniendo especial consideración por los desafíos de probar o aportar la evidencia necesaria para acreditar las distintas formas de discriminación algorítmica. Nuestra hipótesis preliminar es que sólo una vez que comprendamos las distintas formas en que se produce y reproduce la discriminación en la era digital podremos enfrentar los complejos desafíos que tiene la etapa probatoria.

La estructura del artículo es la siguiente: en primer lugar, realizaremos una breve introducción a la más reciente literatura sobre los fundamentos filosóficos del derecho de la antidiscriminación y las implicancias que ello tiene para los desafíos probatorios que deben enfrentar las víctimas; en segundo lugar, describiremos 5 tipos de barreras que deben enfrentar las víctimas de discriminación algorítmica; en tercer lugar, haremos un análisis crítico de un esquema regulatorio emergente (la Ley Local 144 de la Ciudad de Nueva York) que pretende abordar los desafíos probatorios descritos en la sección anterior; finalmente, realizaremos

<sup>3</sup> KITCHIN (2017), p. 15.

<sup>4</sup> AJUNWA (2023), pp. 77-80; ZUIDERVEEN (2020), pp. 1572-1576.

<sup>5</sup> WEF (2022), p. 25.

<sup>6</sup> ALSTON (2018).

<sup>7</sup> NOBLE (2018).

<sup>8</sup> HELLMAN (2020).

<sup>9</sup> FJELD et al. (2020), pp. 4-27.

<sup>10</sup> ZUIDERVEEN (2020), pp. 1576-1582.

algunas conclusiones preliminares y plantearemos futuras preguntas de investigación que surgen para el análisis jurídico.

## 2. El injusto discriminatorio y las implicancias probatorias

La pregunta por los desafíos probatorios que deben enfrentar las víctimas de discriminación algorítmica podría ser abordada, en primer término, desde un punto de vista puramente pragmático, tal como se ha realizado en diversas instancias jurisdiccionales. En efecto, la reflexión puede obedecer a la factibilidad de que la víctima pueda acceder y encontrar evidencia que pruebe la intención, trato o efecto discriminatorio. En el ámbito del derecho europeo de la antidiscriminación, este punto de vista ha llegado incluso a cuestionar las sofisticadas teorías doctrinarias que pretenden distinguir las hipótesis de discriminación: si nos posicionamos desde la perspectiva de la víctima, la pregunta siempre puede ser abordada pragmáticamente, si acaso eso es lo que nos importa al diseñar e implementar un régimen antidiscriminatorio como el europeo<sup>11</sup>. Este punto de vista, sin embargo, evita responder preguntas filosóficas que apuntan a la determinación del injusto discriminatorio o a la injusticia que implica la discriminación, por una parte, y a la necesidad de intervenir con las herramientas propias del derecho, por la otra. En esta sección, abordamos la pregunta por los fundamentos filosóficos del derecho de la antidiscriminación, pues consideramos que ello nos permite abordar los desafíos probatorios que deben enfrentar las víctimas de discriminación algorítmica. En efecto, creemos que el énfasis en la posición de la víctima supone preguntas filosóficas que estimamos necesario desentrañar antes de describir estos desafíos de manera más detallada.

La pregunta por los fundamentos filosóficos del derecho de la antidiscriminación es relativamente reciente. Durante parte importante del siglo XX, y sobre todo desde la publicación y el debate que generó la obra de John Rawls, la pregunta fundamental fue una de justicia distributiva: la pregunta por la justicia, en otras palabras, era principalmente la pregunta por aquello que debemos distribuir de manera igualitaria, tratando de mantener un sano equilibrio entre la preocupación por la igualdad y la libertad, u otros valores asociados<sup>12</sup>. En otras palabras, la pregunta fundamental en la filosofía política era una pregunta sobre la justicia distributiva, sobre aquello que debemos distribuir de manera justa (recursos, libertades, capacidades, oportunidades, bienes u otra “divisa”)<sup>13</sup>. En tal escenario, la discriminación, entendida como un problema entre un agresor y una víctima, era vista como un fenómeno de poco interés para la teoría de la justicia: en la medida en que la discriminación obedece a un problema entre particulares, se pueden aplicar las premisas propias de la justicia comutativa o de las obligaciones que surgen de reparar daños producidos de manera injusta<sup>14</sup>. Desde esta perspectiva filosófica, lo que importaba era determinar la intencionalidad del agente, aquello que residía en la conciencia del agresor, o supletoriamente establecer algún deber de cuidado que eventualmente pudiera haber sido infringido. Ello explica que, por ejemplo, una de las primeras herramientas jurídicas para hacer frente al fenómeno de la discriminación haya sido el derecho penal, como ocurrió con la temprana legislación que tipificaba crímenes raciales en Brasil durante los años 50<sup>15</sup>. Por otra parte, explica el temprano surgimiento, en el derecho anglosajón, de la idea de la discriminación como un daño resarcible, la idea de la discriminación como un *tort*<sup>16</sup>. Sin embargo, este último desarrollo nunca generó mucho debate ni motivó las sofisticadas reflexiones teóricas que se han generado, durante los últimos 50 años, a propósito de la justicia distributiva.

<sup>11</sup> MOREAU (2020), p. 175.

<sup>12</sup> ANDERSON (1999), pp. 287-289.

<sup>13</sup> COHEN (1989), pp. 906-944.

<sup>14</sup> HELLMAN Y MOREAU (2013), pp. 81-86.

<sup>15</sup> CODDOU (2018), pp. 66-67.

<sup>16</sup> MOREAU (2024), pp. 428-429.

Con el desarrollo de investigaciones y experimentos en psicología cognitiva y conductual, se fue lentamente elaborando el concepto de estereotipos implícitos y, posteriormente, el de sesgos inconscientes<sup>17</sup>. A través de múltiples experimentos y estudios empíricos en los más diversos ámbitos – contratación, decisiones judiciales, otorgamiento de créditos, etc– se logró evidenciar cuán contaminado está el razonamiento humano por asociaciones y estereotipos que pasan desapercibidos incluso por actores que deliberadamente intentan evitarlos<sup>18</sup>. Si bien el término “sesgos inconscientes” se acuñó en 1995, las investigaciones se han venido desarrollando y sofisticando de manera progresiva, a tal punto que ya nadie duda que este fenómeno explica gran parte de las decisiones que, de formas individuales o agregadas, terminan desaventajando a ciertos grupos sociales en el acceso a bienes u oportunidades que son fundamentales para la vida de las personas. Estos avances modificaron o desplazaron un poco la pregunta filosófica: si el problema o el origen de la discriminación ya no reside en la mente o en el “espacio cognitivo” del agente, a tal punto que incluso los agentes bienintencionados están contaminados por sesgos, quizás las desventajas que sufren ciertos grupos sociales deben abordarse con otro enfoque.

Un primer cuestionamiento tiene que ver con la pregunta filosófica acerca de la justicia. La discusión filosófica sobre este tema se ha desplazado de la interrogante sobre qué debemos distribuir de manera justa a una que se centra en la justicia de las relaciones sociales<sup>19</sup>. Esta idea de justicia relacional abarca otras dimensiones, como la pregunta por la injusticia de la discriminación. Ahora bien, si nos posicionamos en el lugar de la víctima de la discriminación, los fundamentos filosóficos pueden ser múltiples, lo que ha supuesto un intenso debate que ha permitido el surgimiento del derecho de la antidiscriminación como una disciplina propia<sup>20</sup>. En primer lugar, si desplazamos la preocupación por la injusticia desde el victimario a la víctima, surgen diversas preguntas que vale la pena responder aquí, y que creemos son relevantes para el análisis de los desafíos probatorios de las víctimas de discriminación algorítmica. Desde esta perspectiva, una primera opción supone desechar el concepto de discriminación, pues siempre podríamos dar cuenta de lo que es moralmente problemático sin necesidad de recurrir a la idea de discriminación, ya que podemos encontrar un derecho moral o legal de carácter previo e independiente que está en juego. Por ejemplo, en la denegación de atención de salud a una persona trans, podemos enfocar el problema desde el punto de vista de la infracción al derecho universal a la protección de la salud. Sin embargo, a partir de la masificación de los regímenes jurídicos antidiscriminación y de su desarrollo jurisprudencial —principalmente en Estados Unidos y en el derecho europeo—, la reflexión filosófica mostró que situar la mirada en la perspectiva de la víctima otorga pleno sentido a la pregunta por la fundamentación moral del derecho antidiscriminación. A partir de este debate, surgieron propuestas como las de Sophia Moreau, que conciben la discriminación como una vulneración de las libertades deliberativas<sup>21</sup>. Desde la perspectiva de la víctima, resulta injusto que deba tomar en cuenta características normativamente irrelevantes —como el sexo, la discapacidad o la orientación sexual— en el ejercicio deliberativo que suponen las libertades en sociedades liberales. Bajo esta óptica, la igualdad dejaría de ser el eje central para explicar moralmente la injusticia de la discriminación, pues esta podría entenderse como una violación al derecho a ser tratado como individuo autónomo. En consecuencia, la autonomía y la libertad ofrecerían una base más adecuada para explicar la naturaleza del injusto discriminatorio. Retomaremos esta perspectiva al final del trabajo, ya que creemos que contiene un potencial relevante para abordar los desafíos probatorios de la discriminación algorítmica.

Una teoría alternativa se ha centrado en la dimensión expresiva del acto discriminatorio, entendiendo que este transmite mensajes de hostilidad, animadversión o paternalismo hacia

<sup>17</sup> MOREAU (2020), p. 173.

<sup>18</sup> NORDELL (2021).

<sup>19</sup> CODDOU (2018), p. 121.

<sup>20</sup> MOREAU (2013), pp. 1-2.

<sup>21</sup> MOREAU (2010), p. 147.

determinados grupos<sup>22</sup>. Cuando a ello se suma el efecto de situar a la persona discriminada en una posición o estatus inferior, la discriminación se vuelve injusta y, por tanto, merece reparación. Este enfoque no requiere considerar a la víctima como alguien especialmente dañado o afectado en su esfera individual; basta con un análisis contextual que muestre que queda en una situación inferior para reconocer la incorrección de la conducta. Otras teorías, en cambio, se apoyan en la reflexión filosófica sobre la prioridad, la suficiencia y el mérito dentro de las concepciones de justicia<sup>23</sup>. Desde una teoría moral consecuencialista, que considera que el valor moral se maximiza al mejorar el bienestar de quienes están peor, se justificaría moralmente todo diseño institucional de un régimen antidiscriminatorio que busque mejorar la situación de los grupos históricamente desfavorecidos. Esto se sostiene incluso si las acciones individuales que dieron origen al problema no fueran, en sí mismas, particularmente incorrectas.

Ahora bien, ¿por qué tiene sentido hacer esta introducción sobre los fundamentos filosóficos del derecho de la antidiscriminación en un trabajo sobre los desafíos probatorios de la discriminación algorítmica? En términos generales, si nos enfocamos en el desplazamiento de las preocupaciones filosóficas hacia las víctimas de discriminación algorítmica, es posible pensar de mejor manera en los desafíos probatorios que ellas enfrentan. En primer lugar, y como veremos en la sección siguiente, las víctimas no necesitarían elaborar complejas teorías de la personalidad que permitan atribuir a seres humanos (a los desarrolladores del servicio de reclutamiento y selección de personal de manera automatizada, en el ejemplo) la responsabilidad por daños que se producen por decisiones automatizadas. Además, tampoco necesitarían acceder al algoritmo para demostrar, en algunos casos, por qué ellas fueron tratadas de manera injusta. Bastaría centrarse en el análisis de medidas que, en apariencia neutras, tienen efectos desproporcionados sobre ciertos grupos, cuestión que puede realizarse sin necesidad de probar la intención o de entender de manera detallada cómo se origina o cómo se produce la discriminación en la programación, en el entrenamiento o en la validación de los datos de un algoritmo.

La sección que sigue supone que la reflexión filosófica sobre el derecho de la antidiscriminación ha pavimentado el camino para elaborar una doctrina que permita a las víctimas abordar los desafíos que para ellas supone la prueba de la discriminación algorítmica.

## 2. Las dificultades de probar las diferentes formas de discriminación algorítmica

Los desafíos probatorios para las víctimas de discriminación por parte de procedimientos automatizados de reclutamiento y selección de personal (en adelante, PARS) son múltiples. En esta sección, realizaremos una descripción y análisis de diversas dificultades que podemos identificar a partir de nuestro ejemplo inicial. En términos generales, creemos que estas dificultades, barreras o desafíos pueden entenderse como parte de una reflexión general sobre los principios del derecho probatorio en escenarios donde la asimetría de información es significativa. En este escenario, la necesidad de abordar y superar estos desafíos plantea preguntas más generales sobre los estándares del debido proceso relacionados con la prueba, sobre todo con el derecho a la prueba, entendido este último como el derecho a buscar las fuentes probatorias y articularlas a través de medios probatorios regulados en el proceso, con la finalidad de influir en el resultado de un conflicto jurídico<sup>24</sup>. Al abordar el fenómeno de la discriminación, la pregunta por la “razonable igualdad de posibilidades” que las partes deben tener para influir en el resultado final del proceso se torna especialmente crítica<sup>25</sup>.

### 2.1. *No sabemos si acaso nuestros datos están sometidos a un tratamiento automatizado*

<sup>22</sup> HELLMAN (2008), pp. 13-33.

<sup>23</sup> LIPPERT-RASMUSSEN (2013), pp. 165-170.

<sup>24</sup> ACCATINO (2021), pp. 168-170.

<sup>25</sup> COUTURE (1988), p. 185.

En el caso de nuestro ejemplo inicial, la joven abogada ignora si acaso los datos, personales o no, que ella ha aportado en la plataforma, han sido tratados de manera automatizada. En algunas jurisdicciones, las plataformas que prestan este tipo de servicios están obligadas a realizar notificaciones de carácter explícito y específico acerca del eventual tratamiento automatizado de datos personales<sup>26</sup>. En algunos casos, además, esta notificación debe ir más allá de la mera incorporación de una cláusula en los términos y condiciones de una plataforma. Esta primera barrera es un presupuesto básico, y para derribarla debemos cuestionar la idea de que los sistemas automatizados de decisión son equivalentes a los procesos de decisión humana. En efecto, al cuestionar el “sesgo de automatización”, que supone creer que los resultados o *outputs* de los sistemas automatizados de decisión están libres de sesgos pues en ellos no ha influido el razonamiento humano, podemos empezar a dudar de la “justicia” de las decisiones automatizadas<sup>27</sup>. En nuestro ejemplo, la joven abogada debiera por lo menos tener claridad de que la plataforma recolectaba los datos personales y, que al ponerlos en relación con otros y generar metadatos relevantes para la decisión automatizada en cuestión, recomendaba postulantes a los estudios de abogado para ciertas posiciones.

En el derecho comparado, existen múltiples normas que obligan a los responsables del tratamiento de datos personales a notificar a las personas, de manera explícita, acerca de la recolección de datos personales. Así, por ejemplo, en el caso del Reglamento General de Protección de Datos (en adelante, RGPD), los artículos 13 y 14 establecen obligaciones de notificación cuando se recolecta información directamente de los individuos o a través de terceras personas. Además, el mismo cuerpo normativo incluye una referencia a los procesos o sistemas automatizados de decisión: en términos generales, el artículo 22 supone un derecho de las personas a no ser sujetas a procesos basados únicamente en el tratamiento automatizado de sus datos, con tres excepciones explícitamente contempladas, entre las que se cuenta el consentimiento, siempre y cuando se cumplan con ciertas salvaguardas<sup>28</sup>. Una reconstrucción de estos artículos obligaría a que las obligaciones de notificación no sólo incluyan la recolección de datos personales, sino el tratamiento automatizado de los mismos. De acuerdo con el órgano que controla la interpretación y aplicación coherente del RGPD por parte de los Estados Miembros (en su momento, el Grupo de Trabajo del artículo 29, hoy en día reemplazado por el Comité Europeo de Protección de Datos), si los responsables de un tratamiento de datos quieren invocar el consentimiento para no cumplir con la prohibición de sujetar a las personas a tratamientos automatizados con el objetivo de perfilarlas, entonces deben explicarles de manera muy precisa a qué están consintiendo<sup>29</sup>.

Adicionalmente, la víctima de nuestro ejemplo inicial podría haberse beneficiado de las propuestas que actualmente se discuten en materia de responsabilidad por daños cometidos por los sistemas de IA. En este contexto, la propuesta de Directiva sobre responsabilidad civil por sistemas de IA, presentada por la Comisión Europea en 2022, buscaba facilitar el acceso a la justicia de las personas afectadas por sistemas de IA, incluyendo a víctimas de discriminación algorítmica<sup>30</sup>. Para superar los importantes obstáculos probatorios asociados al funcionamiento opaco y técnico de estos sistemas, la propuesta contemplaba dos herramientas clave: en primer lugar, un mecanismo de acceso a información relevante sobre el sistema de IA, que permitía a los demandantes solicitar a los desarrolladores o usuarios del sistema los datos, parámetros o decisiones técnicas necesarios para sustentar su reclamación, lo que se asemeja al *pretrial discovery* del derecho anglosajón. En segundo lugar, establecía una presunción de causalidad: si

<sup>26</sup> CODDOU Y PADILLA (2024), p. 19.

<sup>27</sup> KUPFER et al. (2023) p. 2.

<sup>28</sup> Esta forma de estructurar la protección ha sido objeto de diversas críticas, en particular por requerir que sea el propio titular del derecho quien active los mecanismos de defensa, y por posibilitar que se vacíe de contenido la protección normativa cuando la intervención humana es mínima o inexistente. PARVIAINEN (2022), pp. 225-248. Por esta razón, algunos autores proponen establecer prohibiciones frente a ciertos tratamientos automatizados de datos, especialmente cuando estos pueden afectar derechos fundamentales. VIOLIER Y FISCHER (2023), pp. 151-159; véase, además, CODDOU Y PADILLA (2024), p. 18.

<sup>29</sup> ARTICLE 29 DATA PROTECTION WORKING PARTY (2018), para. 13.

<sup>30</sup> COMISIÓN EUROPEA (2022), p. 2.

la víctima podía demostrar que se había producido una infracción de una obligación legal relacionada con el uso de IA (por ejemplo, una falla en el deber de diligencia) y que era razonable pensar que tal infracción había causado el daño, entonces se presumía el vínculo causal, desplazando la carga de la prueba al demandado. Estas medidas reconocían la asimetría estructural entre las víctimas y quienes controlan los sistemas de IA, facilitando de forma significativa el ejercicio del derecho a la reparación. Sin embargo, a pesar de estas innovaciones, la Comisión Europea retiró oficialmente la propuesta en febrero de 2025, debido a la falta de consenso político y a cuestionamientos sobre su coherencia con el marco más amplio de regulación de la inteligencia artificial en la Unión Europea, dejando en suspenso la posibilidad de una legislación específica que aborde los daños causados por sistemas algorítmicos<sup>31</sup>.

Además, volviendo a nuestro ejemplo inicial, podemos mencionar que el derecho laboral comparado ha ido paulatinamente creando cláusulas específicas de notificación acerca de la recolección o tratamiento automatizado de datos personales que pueden hacer los empleadores en la etapa precontractual<sup>32</sup>. A nivel estatal, Estados Unidos ha ido paulatinamente creando regulaciones más o menos detalladas sobre la materia, pero todas incluyen cláusulas de notificación en caso de que se estén utilizando procedimientos automatizados<sup>33</sup>. A su vez, el Código del Trabajo francés contiene una sección especial sobre el “reclutamiento” (*recrutement*), que incluye una prohibición para los empleadores de utilizar toda aquella información que no se haya previamente notificado a los candidatos<sup>34</sup>. Además, obliga a los empleadores a cumplir de manera estricta con el principio de finalidad de protección de datos, que exige a los empleadores utilizar la información recolectada de los postulantes solamente para “*evaluar su capacidad para desempeñar el puesto de trabajo ofertado o sus competencias profesionales*”, información que “*debe tener un vínculo directo y necesario con el puesto de trabajo ofertado o con la evaluación de las aptitudes profesionales*”<sup>35</sup>. Para nuestra postulante del ejemplo analizado, no queda claro, por ejemplo, las destrezas que se miden con los juegos o simulaciones o las emociones que se analizan en los análisis de las entrevistas videogramadas. Por otra parte, y con el objeto de habilitar eventuales solicitudes de ajustes razonables, que van más allá de las obligaciones de transparencia algorítmica usualmente aplicables en estos casos, los candidatos deben estar “*informados expresamente, con carácter previo a su realización, de los métodos y técnicas de ayuda a la contratación que se utilicen a su respecto*”, los que a su vez deben ser “*relevantes para el propósito previsto*”<sup>36</sup>.

Esta normativa particular ha sido destacada por ir más allá del marco normativo que actualmente ofrecen los más sofisticados regímenes de protección de datos, que suelen reducirse a la protección de datos personales o sensibles, y que permiten su tratamiento en caso de existir consentimiento. En efecto, parte importante de la discriminación algorítmica en PARS se produce no tanto por el tratamiento de datos personales, obtenidos a través del consentimiento de los postulantes, sino de inferencias potencialmente inexactas o dañinas para personas pertenecientes a ciertos grupos, inferencias que se hacen a partir de la agrupación de datos anónimos<sup>37</sup>. En tal sentido, el deber de notificación que deben cumplir los controladores de datos debe incluir el modo y las finalidades con que se utilizarán los datos recolectados, incluyendo salvaguardas o garantías para la protección del principio de finalidad.

## 2.2. No tenemos una explicación de la forma en que se ha adoptado la decisión

Aun cuando nuestra joven abogada esté al tanto de que la plataforma incluía un sistema automatizado de decisiones, existe una segunda barrera que es crucial analizar si acaso existe la

<sup>31</sup> ANDREWS (2025).

<sup>32</sup> Para una explicación más detallada, véase CODDOU Y PADILLA (2024), pp. 17-19.

<sup>33</sup> ANDREWS (2024).

<sup>34</sup> Code du Travail (Francia), de 2017, Art. L. 1221-9.

<sup>35</sup> Code du Travail (Francia), de 2017, Art. L. 1221-6.

<sup>36</sup> Code du Travail (Francia), de 2017, Art. L. 1221-8.

<sup>37</sup> VILJOEN (2021), pp. 573-654.

posibilidad de encontrar medios de prueba adecuados para el caso. En efecto, nuestra abogada ni siquiera es capaz de elaborar una reconstrucción del caso que le permita entender por qué ella podría ser víctima de una discriminación, y sólo cuenta con algunas intuiciones que la llevan a ser motivo de burla para sus amigas, que consideran exagerada su reacción.

El mismo artículo 22 del RGPD que citamos antes, permite a los controladores de datos hacer perfilamiento de personas en base al tratamiento automatizado de datos personales, siempre y cuando invoquen alguna de las causales que lo admiten (entre otras, el consentimiento del titular, o que sea necesario para la ejecución de un contrato), y cumpliendo con ciertas salvaguardas. En particular, el RGPD obliga al responsable del tratamiento a adoptar “*las medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado, como mínimo el derecho a obtener intervención humana por parte del responsable, a expresar su punto de vista y a impugnar la decisión*”<sup>38</sup>. Según la práctica jurisprudencial y la doctrina, estas salvaguardas pueden interpretarse como parte de lo que se ha considerado como un “debido proceso algorítmico”, que deriva en ciertos derechos que son instrumentales para la protección de otros<sup>39</sup>. Así, el artículo 22.3 del RGPD derivaría en un derecho a solicitar una intervención humana que sea significativa y no un mero trámite burocrático de un funcionario poniendo su firma, en un derecho a obtener una explicación de cómo funciona el sistema automatizado de decisión y cómo se ponderaron los factores o criterios en juego en un caso puntual, y en un derecho a objetar o a impugnar la decisión automatizada<sup>40</sup>.

Si bien en ningún artículo del RGPD se menciona explícitamente el derecho a la explicación, tanto la práctica institucional como la doctrina han considerado que se trata de un derecho fundamental del derecho de la UE<sup>41</sup>. En efecto, de acuerdo con el Recital 71 del RGPD – que según el derecho europeo constituyen instancias que, aunque no jurídicamente vinculantes, son una interpretación autoritativa del derecho europeo que proveen de contexto y permiten comprender el propósito de las normas-, las personas tienen el derecho de conocer una explicación de la decisión individual, pues esa explicación es necesaria para que esa persona pueda ejercer otros derechos – e.g., a impugnar la decisión, a expresar su punto de vista, a la no discriminación – que están explícitamente enumerados en el RGPD. Todos estos derechos, si bien están conectados con el principio fundamental de la transparencia algorítmica, suponen diversas implicancias de acuerdo con las partes o actores involucrados. En el caso de las personas físicas, como señalan diversos autores, y de acuerdo con las directrices elaboradas por el Grupo de Trabajo, el derecho a la explicación supone ser informado acerca de las categorías de datos usadas en el proceso automatizado así como el modo en que los distintos tipos de datos incidieron o fueron relevantes en la decisión individual<sup>42</sup>. Del mismo modo, estas directrices señalan que el derecho a la explicación supone conocer los “*factores que se toman en cuenta para el proceso de toma de decisiones, y (...) su respectivo peso a nivel agregado*”<sup>43</sup>, a conocer cómo se construye un perfil utilizado en la toma de decisiones algorítmicas, “*incluidas las estadísticas utilizadas en el análisis*” y las fuentes de los datos que usaron<sup>44</sup>. El derecho a obtener una explicación, en este sentido, no supone un acceso al código ni a complejas operaciones matemáticas, sino a entender la “lógica” del sistema y el modo en que la misma se reflejó en la decisión individual. Así, el derecho a la explicación no tendría por qué verse afectado con la eventual invocación del secreto industrial o de la propiedad intelectual de la empresa o plataforma sobre el algoritmo en cuestión. Para Veale y Edwards, el tipo de información que debería seguirse de este derecho supone la entrega de información no sólo acerca del modelo (los parámetros de entrenamiento, los datos que fueron usados como insumos, los promedios que se van generando, el modo en que los insumos se transforman en resultados, la forma en

<sup>38</sup> RGPD, artículo 22.3.

<sup>39</sup> KAMINSKI (2021), pp. 198-204.

<sup>40</sup> CONTRERAS Y VIOLLIER (en prensa).

<sup>41</sup> KAMINSKI (2021), p. 204.

<sup>42</sup> ARTICLE 29 DATA PROTECTION WORKING PARTY (2018), para. 31.

<sup>43</sup> ARTICLE 29 DATA PROTECTION WORKING PARTY (2018), para. 27.

<sup>44</sup> ARTICLE 29 DATA PROTECTION WORKING PARTY (2018), para. 31.

que el modelo fue entrenado, etc.) sino de información acerca de la decisión individual, lo que implica un análisis de contrafactuals, de cuáles serían los casos más cercanos a las características del individuo en cuestión, o de bases de datos de personas incluidas y excluidas de una decisión desagregadas demográficamente de acuerdo con las características protegidas por el derecho de la antidiscriminación<sup>45</sup>.

En el caso de nuestra abogada del ejemplo inicial, es más o menos clara la indefensión en que queda, que se explica por la radical asimetría de información, que le impide a la persona siquiera tener acceso al modo en que funciona el sistema. Incluso si la abogada hubiera sido notificada debidamente de que sus datos personales serían procesados por un sistema automatizado, y se le hubiese explicado la “lógica” de su funcionamiento, igualmente debería tener acceso a la forma en que el sistema llegó a la decisión específica en su caso. Además, y como veremos después, la víctima debería tener acceso a bases de datos que incorporen a todas las personas incluidas y excluidas de los algoritmos de recomendación que la plataforma utiliza para hacer el “match” entre estudios de abogados y candidatos, bases de datos que deberían estar desagregadas considerando las categorías protegidas tradicionalmente incluidas en los regímenes antidiscriminación, incluyendo el chileno.

### *2.3. La explicación puede no ser suficiente sin un acceso significativo a los algoritmos*

Hasta el momento, todas las barreras que debería sortear nuestra víctima del ejemplo inicial pueden ser abordadas desde la perspectiva del principio de transparencia algorítmica, que obligaría a notificar a las personas de que sus datos están siendo tratados de manera automatizada, y a tener una explicación acerca del funcionamiento del sistema o del proceso automatizado y del modo en que se ha arribado a una decisión individual en su caso, que en el ejemplo es haberla excluido de la posibilidad de acceder a ciertas posiciones que estaban disponibles en la plataforma. En este último caso, además, la abogada debe tener presente que sus solicitudes de acceder a una explicación pueden implicar una infracción a los secretos comerciales o industriales que protegen a la plataforma. Algunos académicos han planteado su temor a que, en la práctica, las empresas pudieran eludir los requisitos de transparencia del RGPD alegando la necesidad de mantener el “secreto corporativo”<sup>46</sup>. Sin embargo, las directrices citadas antes explican que si bien existe “cierta protección” contra la revelación de secretos comerciales, las empresas “*no pueden confiar en la protección de sus secretos comerciales como excusa para negar el acceso o negarse a proporcionar información*”<sup>47</sup>. Ello, por supuesto, no supone un desconocimiento de la propiedad intelectual o del secreto comercial que pueda estar involucrado en el caso concreto, cuestión que ha quedado clara en el Recital 63, a propósito del derecho de acceder a la información pertinente acerca del tratamiento de datos personales. Como señala este Recital, “[e]ste derecho no debe afectar negativamente a los derechos y libertades de terceros, incluidos los secretos comerciales o la propiedad intelectual y, en particular, los derechos de propiedad intelectual que protegen programas informáticos. No obstante, estas consideraciones no deben tener como resultado la negativa a prestar toda la información al interesado. Si trata una gran cantidad de información relativa al interesado, el responsable del tratamiento debe estar facultado para solicitar que, antes de facilitarse la información, el interesado especifique la información o actividades de tratamiento a que se refiere la solicitud”<sup>48</sup>.

En términos generales, y más allá de estas pretensiones normativas por lograr que la transparencia algorítmica no implique la desprotección de otros intereses privados, el escollo no es menor. En primer lugar, y dependiendo de la definición de algoritmo que adoptemos, estos podrían ser protegidos por el derecho de autor si acaso son consideradas obras creadas por

<sup>45</sup> VEALE Y EDWARDS (2017), pp. 55-58.

<sup>46</sup> MALGIERI (2019), p. 25.

<sup>47</sup> ARTICLE 29 DATA PROTECTION WORKING PARTY (2018), para. 2.

<sup>48</sup> RGPD, Recital 63.

seres humanos. Si bien en principio los algoritmos son creados, desarrollados y entrenados por seres humanos (generalmente, de las ciencias de la computación), existe un sinnúmero de algoritmos creados de manera enteramente artificial<sup>49</sup>. En segundo lugar, si los algoritmos fuesen considerados como invenciones, podrían ser patentables por el derecho de la propiedad industrial. Sin embargo, en varias legislaciones, incluyendo la chilena, los algoritmos serían entidades abstractas, “*un conjunto de operaciones que se pretende realizar para conseguir un resultado concreto o resolver un problema*” equiparable a los modelos matemáticos, que no serían patentables<sup>50</sup>. Más allá de este debate, es efectivo que las empresas o plataformas podrían argüir que los algoritmos están protegidos por el derecho de propiedad intelectual o industrial en la medida que forman parte de programas computacionales, o en la medida en que se argumente que constituyen algo más que un modelo matemático abstracto, y en tal caso registrables o patentables, convirtiéndose en barreras para las víctimas de discriminación algorítmica.

#### 2.4. La prueba de la discriminación algorítmica directa

Aunque nuestra víctima del ejemplo inicial hubiera pasado todas las barreras explicadas anteriormente, y ya cuente con la información necesaria que le permita iniciar un reclamo o cuestionar una decisión de contratación, nos encontramos con un nuevo obstáculo. En efecto, nuestra víctima ha debido sobrepassar no sólo la asimetría de información que se produce al intentar probar la discriminación en etapa precontractual, sino por la radical asimetría de información que se produce debido a la intermediación algorítmica del proceso de reclutamiento. En esta etapa, la víctima se enfrenta a la carencia de los conocimientos técnicos y los recursos necesarios para evaluar la imparcialidad de estas herramientas o para presentar acciones de reclamación<sup>51</sup>. En términos generales, y como señala Coloma, en etapas prejudiciales “[l]a pretensión de la parte demandante o acusadora obedece (...) a una elección en la que se han de tener en cuenta variables como pruebas potencialmente disponibles, prácticas interpretativas asentadas y previsión de los costes de litigación”<sup>52</sup>.

En los casos de discriminación algorítmica, las víctimas carecen del conocimiento experto o de los recursos necesarios para tratar de probar una intención discriminatoria, de buscar un comparador real o hipotético que les permita probar la causalidad o, por último, generar evidencia suficiente para probar el impacto discriminatorio<sup>53</sup>. Para el análisis de este desafío, entonces, es importante distinguir entre la discriminación directa e indirecta, cuestión que ha estado en el centro del debate doctrinario desde el surgimiento mismo del derecho de la antidiscriminación como una disciplina independiente<sup>54</sup>.

De acuerdo con una definición más o menos estándar, tal como se incluye en diversos regímenes antidiscriminación, se considera discriminación directa la situación en que se encuentra una persona que sea, haya sido o pudiera ser tratada, en atención a algunos de los criterios o categorías tradicionalmente protegidas por el derecho, de manera menos favorable que otra en situación comparable. Esta definición asume varias cuestiones: en primer lugar, que debe invocarse un criterio comparador que causa la diferencia de trato; en segundo lugar, que debe probarse el trato menos favorable en relación a algo, sea un derecho, un beneficio o un interés de otro tipo; en tercer lugar, que la explicación causal que existiría entre el criterio o

<sup>49</sup> NAUGHTON (2022).

<sup>50</sup> AZUAJE-PIRELA Y FINOL (2022), p. 162.

<sup>51</sup> SERRANO (2021), p. 191.

<sup>52</sup> COLOMA (2017), p. 36.

<sup>53</sup> GINÈS (2021), p. 311; CODDOU Y PADILLA (2024), p. 19.

<sup>54</sup> COLLINS Y KHAITAN (2018), p. 24; HELLMANN Y MOREAU (2013), pp. 1-4. Si bien existen otras categorías doctrinarias que han llegado a ser reconocidas por diversos arreglos institucionales, como el caso de la discriminación por asociación, por error o interseccional, éstas se relacionan con la función que cumplen las categorías protegidas en la reconstrucción normativa del derecho antidiscriminación. Más allá de que esta función puede tener implicancias probatorias, reduciremos nuestro análisis a las dos principales categorías doctrinarias reconocidas por el derecho comparado de la antidiscriminación y relacionaremos estas otras categorías con las hipótesis de discriminación directa o indirecta. Agradecemos a uno de los árbitros por este comentario.

categoría y el trato menos favorable puede hacerse de manera retroactiva (es decir, que la víctima invoque un caso concreto en que una persona en situación comparable haya sido tratada de manera más favorable) o a través de un comparador hipotético<sup>55</sup>. Esta definición, por tanto, cuestiona la idea de que la discriminación directa supone la prueba de la intención discriminatoria, cuestión que ha quedado atrás luego de que la preocupación por el injusto discriminadorio se desplazó desde el autor hacia la víctima de la discriminación. Más allá de que en algunas jurisdicciones y, por sobre todo, en la práctica jurisprudencial, la intención discriminatoria siga siendo el criterio determinante a la hora de acreditar una hipótesis de discriminación<sup>56</sup>, existe un consenso en el derecho comparado de la antidiscriminación de que la discriminación directa puede producirse de formas no intencionales. Además, este cuestionamiento supone que se pueden probar casos de discriminación directa aunque el actor no haya realizado mención alguna a una categoría o criterio protegido en la realización de la oferta, en la decisión de denegar un crédito o en el rechazo de una solicitud de arriendo<sup>57</sup>.

En este sentido, salvo en los casos en que sea posible probar la intención, a través de alguna inferencia que puede realizarse, por ejemplo, por la mención explícita a alguna categoría protegida en el acto que se cuestiona como discriminatorio, la prueba de la discriminación directa queda entregada al test probatorio de la desigualdad de trato en relación a personas similarmente situadas (*sine qua non*). De este modo, se requeriría probar no sólo que una de las categorías protegidas fue considerada entre las razones para una exclusión o rechazo de una oferta o de un beneficio, sino la razón determinante. Para otros, sin embargo, este test supone una cuestión puramente causal que se puede hacer de manera empírica o hipotética, sin necesidad alguna de involucrarse con el razonamiento o la deliberación del actor<sup>58</sup>. Por otra parte, y tal como señala Kelly-Lyth, lograr acceso a los datos de entrenamiento o de implementación de un algoritmo (por ejemplo, a través de tablas de confusión) nos permitiría tener condiciones ideales para realizar este tipo de análisis, sobre todo en contraste con la evaluación del razonamiento humano<sup>59</sup>.

En el caso de nuestra joven abogada, quizás lo más fácil sea probar un caso de discriminación directa si acaso se consideró el sexo entre los parámetros con que se diseñó la plataforma automatizada de reclutamiento y selección. Por ejemplo, si la plataforma tomó como base para la predicción de la productividad de los candidatos a la lista de los mejores abogados de la plaza o del estudio en cuestión, y resulta que todos ellos son hombres, entonces puede que se haya decidido por optar diseñar el algoritmo con esa instrucción discriminatoria. Sin embargo, la mayor parte de las veces estos criterios no operan de manera explícita o aparente como un parámetro de los algoritmos. Los regímenes comparados del derecho de la antidiscriminación han reaccionado creando categorías como la discriminación por asociación o la discriminación a través de *proxies*<sup>60</sup>. En el caso de nuestro ejemplo, la plataforma puede haber estado diseñada con parámetros que son discriminatorios de manera explícita, pero las más de las veces se trataría de criterios (*proxies*) que están inherentemente relacionados con alguna categoría protegida y que se van aprendiendo a medida que el algoritmo va siendo entrenado y testeado con postulaciones de la vida real<sup>61</sup>. Aun cuando la plataforma haya sido diseñada evitando incorporar el sexo entre los parámetros de selección, el algoritmo puede haber aprendido, a través del entrenamiento y retroalimentación con los datos de diversos

<sup>55</sup> GOLDBERG (2010), p. 805.

<sup>56</sup> MUÑOZ (2015), p. 161.

<sup>57</sup> Para un análisis que distingue la discriminación directa e indirecta sin depender de lo explícito o encubierto del criterio o categoría protegida, véase Tribunal de Justicia de la UE, *CHEZ Razpredelenie Bulgaria AD y Komisia za zashtita ot diskriminatsia*, C-83/14, 16 de Julio de 2015.

<sup>58</sup> Véase la discusión acerca del famoso caso de *James v. Eastleigh*, pronunciado por la Cámara de los Lores, en BAMFORTH et al. (2008), pp. 264-268.

<sup>59</sup> KELLY-LYTH (2023), p. 160.

<sup>60</sup> En estos casos, para evaluar si estamos ante un caso de discriminación directa, debemos analizar el grado de coincidencia entre el proxy y la categoría protegida. De este modo, antes de que se reconociera el embarazo como categoría protegida, se reconocía que estos casos eran hipótesis de discriminación por sexo.

<sup>61</sup> KELLY-LYTH (2021), p. 904. En efecto, los casos de discriminación inherente (en que un proxy está evidentemente relacionado con una categoría protegida) son un tipo de discriminación directa. KELLY-LYTH (2023), p. 160.

postulantes, que algunos criterios que están relacionados con la productividad están asociados, a su vez, con características que se observan sólo en hombres, como la práctica de ciertos deportes o la estadía en trabajos previos.

Quizás, en el caso de nuestra abogada, la plataforma reclutante trabajó con los estudios de abogados que contrataron el servicio para definir criterios o calificaciones como determinantes para la tarea de filtrar postulantes. Sin embargo, el principal problema se debe a que los algoritmos, por regla general, no identifican relaciones de causalidad, sino de correlación estadística<sup>62</sup>. De este modo, si la mayoría de los “buenos abogados” son hombres, el algoritmo aprenderá, a través de su entrenamiento, que determinados rasgos masculinos (la práctica de ciertos deportes o el uso de cierto lenguaje) son predictivos de buenos futuros empleados<sup>63</sup>. Esto genera un círculo vicioso: como el algoritmo está entrenado de esa manera, existirá más probabilidad de que abogados hombres sean seleccionados, amplificando la probabilidad de que, en el futuro, mujeres similarmente situadas corran el riesgo de quedar excluidas de la selección por la plataforma de reclutamiento<sup>64</sup>.

Lo que habría que probar, en este caso, es que el entrenamiento o retroalimentación de los datos están contaminados por sesgos de género en diversas etapas del ciclo de vida de un algoritmo. En algunos casos, el origen de la discriminación algorítmica estará radicado en el sesgo muestral o estadístico, que en algunos grandes empleadores o reclutadores puede llegar a ser significativo. Así, si la base de datos que permite a las empresas construir los criterios de “fit cultural” está compuesta de una mayoría significativa de hombres, la prueba de la discriminación algorítmica podría requerir revelar las fuentes o bases de datos utilizadas para el entrenamiento o las que (retro)alimentan los procesos de implementación. En otros casos, lo que la víctima debería probar es qué es lo que cuenta como un “buen trabajador” o como un trabajador “productivo”: si un sistema de selección automatizado se basa en indicadores como la duración en empleos anteriores, en lugar de criterios individuales como evaluaciones personales, es probable que se generen efectos adversos sobre ciertos grupos protegidos. Por ejemplo, las mujeres suelen tener trayectorias laborales más interrumpidas debido a responsabilidades de cuidado, lo que puede afectar negativamente la evaluación de sus currículums por parte de sistemas automatizados de filtrado (*applicant tracking systems*)<sup>65</sup>. Además, si se da prioridad a las evaluaciones de jefaturas pasadas para definir quién es un “buen trabajador”, existe el riesgo de que se reproduzcan sesgos humanos presentes en esos datos. En este caso, el problema no radica tanto en un error estadístico, sino en cómo se recolectan y analizan los datos, lo que exige demostrar de qué forma los prejuicios humanos influyeron en el proceso. En general, los *proxies* utilizados en estos sistemas no necesariamente se correlacionan de forma exacta con categorías protegidas, lo que puede excluirlos de los conceptos tradicionales de discriminación por asociación o por *proxy*. No obstante, tanto la jurisprudencia europea como la británica han sido flexibles al respecto y no exigen una correspondencia exacta entre el indicador utilizado y la categoría protegida para reconocer la existencia de discriminación<sup>66</sup>. Para la prueba de la discriminación directa, recordemos, basta que existan buenas razones para probar que ha existido una desigualdad de trato con personas similarmente situadas, y esa desigualdad de trato puede ser determinada a través del mayor riesgo que sufren las mujeres de ser excluidas de la selección o reclutamiento para un determinado trabajo<sup>67</sup>.

La pregunta por la distinción entre casos de discriminación directa e indirecta no es sólo filosófica o doctrinaria, sino que tiene efectos prácticos en diversos regímenes jurídicos. En algunos casos, la discriminación directa es considerada de tal gravedad que no se admiten

<sup>62</sup> XENIDIS (2020), p. 749.

<sup>63</sup> KELLY-LYTH (2021), p. 903.

<sup>64</sup> ADAMS-PRASSL et al. (2023), p. 8.

<sup>65</sup> BAROCAS Y SELBST (2016), pp. 679-680.

<sup>66</sup> Tribunal de Justicia de la UE, *Tadao Maruko contra Versorgungsanstalt der deutschen Bühnen*, C-267/06, 1 de Abril de 2008; Reino Unido, Corte Suprema, *R (on the application of Coll) v Secretary of State for Justice* (2017) UKSC 40.

<sup>67</sup> ADAMS-PRASSL et al. (2023), p. 19.

defensas en contrario, tal como se deriva de la Ley de Igualdad del Reino Unido<sup>68</sup>. Lo mismo podríamos derivar de una lectura del artículo 2º inciso sexto del Código del Trabajo chileno, que señala que son actos de discriminación las ofertas de trabajo efectuadas por un empleador, directamente o a través de terceros y por cualquier medio, que señalen como un requisito para postular a ellas alguna de las categorías protegidas que se mencionan, con un carácter abierto, en el mismo artículo<sup>69</sup>.

## 2.5. *La prueba de la discriminación algorítmica indirecta*

Más allá de las hipótesis de discriminación algorítmica directa que analizamos recién, una parte importante de la literatura académica acerca de la prueba de la discriminación algorítmica se enfoca en los casos de discriminación indirecta<sup>70</sup>. El razonamiento se puede describir del siguiente modo: considerando la dificultad de probar una intención discriminatoria, de encontrarse con casos en que se ha utilizado de manera explícita una categoría protegida en el diseño de un algoritmo, o de escudriñar entre las razones o criterios que ha considerado el algoritmo para evaluar si acaso alguna categoría protegida operó como la razón determinante, es plausible pensar que la discriminación algorítmica puede ser mejor abordada a través de la categoría de la discriminación indirecta. Nuevamente, de acuerdo con una definición estándar, la discriminación indirecta supone casos en que una disposición, criterio o medida en apariencia neutra genera un impacto desproporcionado en personas que poseen u observan algunas de las categorías protegidas<sup>71</sup>. De ahí que, desde la perspectiva de las víctimas, y considerando que estamos frente a verdaderas “cajas negras”, cuyo modelo o decisión individual no podemos explicar, sea relativamente obvio considerar la discriminación indirecta como un punto de partida<sup>72</sup>.

Aún más problemático es cuando los criterios utilizados para evaluar postulantes no guardan una relación real con las competencias o habilidades laborales requeridas, sino que se basan en correlaciones espurias o sin una justificación aparente. Un caso emblemático ilustra esta situación: un sistema identificó el nombre “Jared” y la práctica del deporte “Lacrosse” como los mejores predictores de buenos candidatos<sup>73</sup>. En estos escenarios, se utilizan criterios de perfilamiento que no están vinculados directamente a categorías protegidas, pero que resultan arbitrarios, ya que las personas no tienen control sobre esos aspectos<sup>74</sup>. No obstante, incluso en estos casos sería posible plantear una hipótesis de discriminación, si se demuestra que existe una asociación significativa entre esas variables aparentemente irrelevantes y una categoría protegida. Esta lógica es la que subyace en muchos casos de discriminación indirecta abordados a través del concepto de discriminación por asociación, donde la validez del reclamo depende de la fuerza o intensidad de dicha correlación.

Diversos arreglos institucionales han tratado de generar presunciones, estándares o métricas de la discriminación indirecta (el término utilizado en Europa) o del impacto desproporcionado (el término utilizado en Estados Unidos). En Estados Unidos, la prueba de un impacto desproporcionado asociado a decisiones adoptadas por el empleador ha dado lugar a controversias centradas en peritajes económicos, los cuales buscan demostrar desviaciones estadísticamente significativas o estimaciones mediante regresiones lineales que permitan aislar las variables responsables de dicho impacto. Un ejemplo emblemático de este enfoque fue el caso impulsado por Betty Dukes, quien, a través de una acción colectiva por discriminación de

<sup>68</sup> En el caso de la ley antidiscriminación del Reino Unido, los casos de discriminación directa no admiten excepción, salvo en relación con el criterio de la edad, siempre y cuando se cumplan ciertas salvaguardas. *Equality Act* (2010), sección 13. En otros casos, además, se pueden negociar excepciones concretas y detalladas, como consta en el anexo de aquella regulación.

<sup>69</sup> Véase CODDOU Y PADILLA (2024), pp. 25-26.

<sup>70</sup> XENIDIS (2020), p. 747.

<sup>71</sup> XENIDIS (2023), p. 384.

<sup>72</sup> ADAMS-PRASSL et al. (2023), pp. 4-7.

<sup>73</sup> BOGEN (2019).

<sup>74</sup> WACHTER (2022), pp. 5-7.

género, demandó a la empresa Walmart<sup>75</sup>. Además, el derecho laboral estadounidense es conocido por la implementación de test psicométricos que permitan probar o, al menos, presumir la existencia de un sesgo desfavorable hacia ciertos grupos en decisiones de contratación<sup>76</sup>. En Estados Unidos, la Comisión para la Igualdad de Oportunidades en el Empleo (EEOC) ha utilizado, desde 1978, la denominada “regla de los cuatro quintos”, con el objeto identificar posibles impactos adversos en los procesos de selección laboral<sup>77</sup>. La directriz establece que si la tasa de contratación de un grupo protegido es inferior al 80% de la tasa del grupo con mayor éxito, podría existir una presunción de discriminación indirecta. Por ejemplo, si en un proceso de selección el 60% de los candidatos blancos son contratados, pero solo el 20% de los candidatos afroamericanos lo son, la tasa relativa ( $20/60 = 0.33$ ) está por debajo del umbral del 0,8, lo que sugiere un impacto adverso hacia los candidatos afroamericanos<sup>78</sup>. En tales casos, corresponde al empleador demostrar que el criterio de selección utilizado es válido y necesario para el desempeño del puesto.

De todos modos, estos intentos por facilitar o alivianar la carga probatoria pueden resultar inútiles si no existen obligaciones de transparencia activa que permitan aminorar el costo y el tiempo que supone el ejercicio individual de las acciones de reclamación<sup>79</sup>. Por ello, y para que este tipo de reglas de presunción funcionen adecuadamente, se requiere contar con obligaciones de registro desagregado de datos y de transparencia algorítmica. En algunas jurisdicciones, además, se permite el ejercicio de acciones colectivas o de interés público, otorgando legitimidad activa a organizaciones sindicales o autoridades que representan el interés público comprometido, como instituciones nacionales de derechos humanos, defensorías ciudadanas, inspecciones del trabajo u otras agencias de protección de derechos<sup>80</sup>. Con información desagregada y actualizada, incluyendo tanto a los individuos seleccionados como no seleccionados en procesos de reclutamiento, se pueden adoptar medidas para la corrección de sesgos, que en algunos casos pueden llegar a ser obligatorias. Para nosotros, una reconstrucción normativa del artículo 10.5 de la Ley de la IA de la UE, en concordancia con el artículo 9.2 letra g) del RGPD y con el principio de no discriminación (que ha adquirido un carácter fundamental en el derecho de la UE), no sólo permite, sino que obliga a los responsables de los sistemas considerados como de “alto riesgo” a recolectar datos sensibles que de otro modo estarían prohibidos (como como datos de salud o de origen étnico) en la medida en que sean estrictamente necesarios para el seguimiento, detección y corrección de sesgos<sup>81</sup>. Como parece evidente, este tipo de medidas correctivas debe complementarse con medidas que limiten la reutilización o transferencia de datos y que permitan la adopción de técnicas de preservación de la privacidad, incluyendo el uso de datos sintéticos o simulados (por ejemplo, para disminuir el sesgo estadístico de una base de datos), la anonimización o la pseudonimización. Este tipo de medidas parecen ser determinantes a la hora de facilitar procesos de evaluación de impacto (*ex ante*), que son obligatorias, o de auditorías algorítmicas, que pueden realizarse *a posteriori*, como veremos en la siguiente sección<sup>82</sup>. Por último, además del acceso a información relevante o del establecimiento de medidas preventivas de debida diligencia o de cumplimiento de estándares sectoriales, se requiere una regla que permita el desplazamiento de la carga de la prueba, cuestión que ha intentado instalarse a través de la adaptación de las normas sobre responsabilidad civil a los daños cometidos por sistemas de IA.

<sup>75</sup> Estados Unidos: Wal-Mart v. Dukes, 10-277 U.S. (2011).

<sup>76</sup> STETZ (2022), p. 68.

<sup>77</sup> EEOC, Uniform Guidelines on Employee Selection Procedures, de 1978. Tal como su nombre señala, se trata de directrices (*guidelines*) –es decir, no son leyes ni reglamentos– que se dictaron para orientar a empleadores y otras entidades en el cumplimiento del derecho. En el caso *Griggs v. Duke Power* (1971), la Corte Suprema de Estados Unidos estableció que los tribunales deben otorgarles una “gran deferencia”. STETZ (2022), p. 69.

<sup>78</sup> STETZ (2022), pp. 67-74.

<sup>79</sup> Los párrafos finales de esta sección son una adaptación de CODDOU Y PADILLA (2024), pp. 21-22.

<sup>80</sup> ADAMS-PRASSL et al. (2023), pp. 124 – 151.

<sup>81</sup> En efecto, si bien el tenor literal del artículo 10.5 de la EU AI Act sostiene que se trata de una permisión, una reconstrucción normativa más integral sugiere que las infracciones al derecho o principio de no discriminación, que es fundamental para el derecho europeo, se producen también por omisiones o infracciones a deberes de cuidado razonables. En el mismo sentido, véase CALVI (2023), pp. 23-25.

<sup>82</sup> KIM (2022), p. 1550.

En la jurisprudencia del derecho europeo sobre discriminación indirecta, el criterio utilizado contrasta con métricas de equidad estadística. En concreto, la Corte de Justicia de la UE ha ido consolidando el criterio de la paridad demográfica, que consiste básicamente en intensificar el escrutinio de control judicial allí donde hay una minoría significativa de personas del grupo desventajado en posiciones de beneficio<sup>83</sup>. A propósito de la discriminación algorítmica, la pregunta sería entonces por la combinación de métricas cuantificables que puedan complementarse con análisis contextuales propios del contexto de adjudicación. De todos modos, para autoras como Xenidis, la jurisprudencia europea ha errado en adaptar de una manera puramente técnica y neutral equivalencias entre los problemas de equidad que surgen en las ciencias de la computación y el derecho antidiscriminación. Para Raphaële Xenidis, es importante salir tanto de la trampa del encuadre (*framing trap*) como de la trampa del formalismo (*formalism trap*), que nos impiden salir de un análisis meramente técnico del fenómeno de la discriminación algorítmica, que sería “*no solo el producto de algoritmos sesgados*”, sino que sería “*coproducida en la intersección de prácticas epistémicas, sociales y técnicas, es decir, en la formulación de problemas, el despliegue tecnológico, las lógicas organizacionales y los procesos humanos*”<sup>84</sup>. Más aun, como ha sostenido recientemente Pauline Kim, si los tests de discriminación algorítmica derivan en métricas de equidad estadística desprovistas de contexto, podemos terminar anclados a un reduccionismo estadístico, olvidando la posición institucional privilegiada que tiene el empleador, quien debe tener la obligación de demostrar la validez sustantiva y técnica de sus herramientas de reclutamiento y selección<sup>85</sup>. De este modo, la carga del empleador no se debería reducir a mostrar el cumplimiento de cierto criterio de equidad estadística, sino a demostrar que la tecnología en cuestión se diseñó utilizando datos precisos, representativos e imparciales, y que es pertinente y adecuada para medir las habilidades y capacidades relevantes para el trabajo en cuestión<sup>86</sup>. Para algunos/as investigadores/as, los empleadores que utilizan estas tecnologías deberían además estar disponibles para que se realicen auditorías independientes o que sus procesos estén respaldados por evidencia o estudios que admitan la evaluación de pares, cuestión que no ha sucedido con ninguna de las grandes compañías tecnológicas que usan PARS<sup>87</sup>. De otro modo, quedarían sujetos a hipótesis de responsabilidad por productos o servicios defectuosos, que no pueden realizar lo que prometen cumplir en su estrategia de comercialización.

El “testeo situacional” es otra de las herramientas probatoria más innovadoras en casos de discriminación, que consiste en evaluar si individuos con características similares, pero pertenecientes a diferentes grupos protegidos, reciben un trato desigual en contextos específicos. Este método busca establecer una relación causal entre una medida o un proceso aparentemente neutral y un impacto desproporcionado en ciertos grupos, difuminando la distinción entre discriminación directa e indirecta. En el contexto de la discriminación algorítmica, aplicar el testeo situacional presenta desafíos particulares. A diferencia de los métodos tradicionales, como el del “cliente misterioso” utilizado en estudios de discriminación contra consumidores, la evaluación de sistemas automatizados requiere enfoques más sofisticados<sup>88</sup>. Por ejemplo, en el caso *Sandvig v. Barr*, investigadores académicos crearon perfiles ficticios para interactuar con procesos automatizados de selección, con el fin de detectar posibles sesgos discriminatorios<sup>89</sup>. El Tribunal de Distrito de Columbia determinó que esta práctica no violaba la Ley de Fraude y Abuso Informático (CFAA), ya que los investigadores accedieron a sitios públicos sin contraseñas, y su conducta no excedió el acceso autorizado. En contraste con métricas cuantitativas como la regla del 80/20, que establece umbrales específicos para identificar disparidades en tasas de selección, el testeo situacional ofrece una perspectiva

<sup>83</sup> WACHTER et al. (2021), pp. 8-9.

<sup>84</sup> XENIDIS (2023), p. 385.

<sup>85</sup> KIM (2017), pp. 189-203.

<sup>86</sup> CODDOU Y PADILLA (2023), p. 22.

<sup>87</sup> NARAYANAN Y KAPOOR (2024), p. 25.

<sup>88</sup> NEVEN et al. (2022).

<sup>89</sup> Estados Unidos, District Court of Columbia, *Sandvig v. Barr* (2020).

más cualitativa y contextualizada. Ambos enfoques son complementarios y, cuando se utilizan conjuntamente, proporcionan una evaluación más integral de posibles prácticas discriminatorias en entornos laborales.

### **3. Una propuesta concreta: el caso de la Ley Local 144 de la Ciudad de Nueva York**

Como hemos visto a lo largo de este trabajo, los desafíos probatorios en casos de discriminación algorítmica son múltiples y complejos, especialmente cuando se enfrentan procesos de toma de decisión automatizada que se presentan como opacos, impersonales y supuestamente neutrales. Estos desafíos van desde la falta de notificación a las personas afectadas sobre el uso de herramientas algorítmicas, hasta la imposibilidad de acceder a explicaciones comprensibles, desagregación de datos, o información que permita establecer relaciones causales, correlaciones o impactos desproporcionados. En este contexto, resulta pertinente examinar cómo ciertos regímenes jurídicos han comenzado a enfrentar estos problemas mediante nuevos esquemas de regulación. Esta sección analiza la Ley Local 144 de la Ciudad de Nueva York (en adelante, LL144NYC) como una propuesta concreta para abordar parte de estas barreras, particularmente en el ámbito del empleo y la contratación automatizada<sup>90</sup>. La inclusión de esta sección permite mostrar cómo los problemas probatorios analizados en este artículo no son solo una cuestión teórica, sino que ya han motivado respuestas regulatorias concretas que ilustran tanto los avances como las limitaciones de dichas soluciones. Además, analizar la LL144NYC funciona como un “caso de estudio” que vincula directamente algunas de las barreras probatorias descritas en las secciones anteriores con intentos normativos de redistribuir la carga de la prueba y reducir la asimetría de información que enfrentan las víctimas.

La LL144NYC, también conocida como la “ley de auditoría de sesgos algorítmicos”, fue adoptada en 2021 por el Concejo Municipal de Nueva York y entró en vigor en julio de 2023. Su elaboración fue impulsada por organizaciones de derechos civiles y, si bien en un principio se focalizaba en los sistemas de contratación pública, derivó en una regulación de las herramientas diseñadas e implementadas por las empresas privadas<sup>91</sup>.

Los objetivos principales de la ley son “*prevenir la discriminación algorítmica en la contratación laboral*”, asegurar que estas tecnologías promuevan la equidad y no reproduzcan sesgos estructurales, y “*crear un marco regulatorio que facilite la rendición de cuentas*”. Para ello, la ley establece varias obligaciones significativas: auditorías de sesgo anuales a cargo de un auditor independiente, que evalúan el impacto de las herramientas automatizadas de decisión de empleo (AEDT) sobre grupos protegidos según raza/etnia y sexo (estas auditorías deben calcular tasas de selección y razones de impacto, y publicarse en el sitio web del empleador); notificación obligatoria a las personas postulantes sobre el uso de AEDT, su funcionamiento general y sus derechos, incluyendo el derecho a solicitar procesos alternativos o ajustes razonables; y, por último, requisitos de transparencia, como la divulgación pública de los resultados de las auditorías, las fuentes de datos utilizadas y los métodos de evaluación aplicados, los cuales deben estar disponibles durante al menos seis meses.

Desde una perspectiva probatoria, la importancia de esta ley radica en que traslada parte de la carga probatoria al empleador, y permite a las personas afectadas acceder a información básica para evaluar si se ha producido o no una forma de discriminación. Así, por ejemplo, las obligaciones de auditoría y transparencia permiten generar evidencia sobre impactos desproporcionados, que pueden constituir indicios suficientes para activar mecanismos legales. Asimismo, el deber de notificación y la posibilidad de solicitar procesos alternativos refuerzan la capacidad de acción de las personas frente a decisiones automatizadas, y habilitan mecanismos de control social y jurídico sobre el funcionamiento de los algoritmos utilizados en el empleo. En este sentido, una opción recomendada por el Center for Democracy and Technology es

<sup>90</sup> NYC CONSUMER AND WORKER PROTECTION (2023).

<sup>91</sup> FUCHS (2023), pp. 209-211.

precisamente que el test automatizado sea opcional, y que los candidatos que no tengan una preferencia sean aleatoriamente asignados entre el sistema automatizado y el alternativo, de manera que se asegure un tratamiento igualitario<sup>92</sup>.

Si bien la Ley Local 144 de Nueva York representa un esfuerzo pionero por regular el uso de herramientas automatizadas en decisiones de empleo, su diseño y aplicación han sido objeto de críticas significativas, que pueden organizarse en tres dimensiones principales: el texto legal, el diseño regulatorio y la implementación práctica.

### 3.1. Críticas al texto legal y a las definiciones normativas

Una de las principales debilidades señaladas es la vaguedad en la redacción del texto legal, especialmente en conceptos clave como “*Herramienta Automatizada de Decisión de Empleo*” (AEDT) y el criterio de si esta “*ayuda sustancialmente o reemplaza*” la toma de decisiones humanas. En otras palabras, la regulación otorga amplia discrecionalidad a los empleadores para determinar si las herramientas que emplean son efectivamente AEDTs bajo los términos de la LL144NYC y, por ende, si deben regirse por ella. Es el propio empleador quien tiene la capacidad de decidir si su herramienta cae bajo el ámbito de aplicación de la norma, lo cual genera el denominado riesgo de cumplimiento nulo (*null compliance*)<sup>93</sup>.

Esta ambigüedad crea lagunas interpretativas que dificultan identificar con precisión qué sistemas deben ser auditados. Además, la ley exige que las auditorías sean realizadas por una entidad “independiente”, pero no establece criterios claros de independencia más allá de la financiera, lo que genera incertidumbre sobre quién califica como auditor válido. En contraste con los criterios de independencia contenidos en regulaciones como la Ley Sarbanes-Oxley de EE.UU. (2002), que prohíbe estrictamente que los auditores independientes presten “*ciertos servicios no relacionados con auditorías*” a sus clientes, muchas de las auditorías de la LL144NYC terminaron siendo realizadas por otros departamentos de la misma empresa o por consultoras que prestaban otros servicios a los empleadores<sup>94</sup>.

En consecuencia, la ley termina socavando su propia finalidad. Al otorgar a las mismas empresas reguladas la discrecionalidad de decidir si sus AEDT se encuentran dentro del ámbito de aplicación, se abre la puerta a que simplemente se excluyan del cumplimiento. Incluso en el caso de que una empresa acepte quedar sujeta a la norma, la obligación de auditoría puede cumplirse de manera meramente formal, recurriendo a un “consultor amigo” o incluso a otro departamento interno de la compañía, lo que anula en los hechos el requisito de independencia. Este diseño debilita la meta central de la ley —garantizar transparencia en el uso de datos— pues permite que las empresas eludan la notificación a los candidatos, les nieguen explicaciones sobre las decisiones tomadas y, en última instancia, impidan el acceso a auditorías imparciales.

### 3.2. Críticas al diseño regulatorio

En cuanto a su arquitectura normativa, se ha cuestionado la falta de umbrales mínimos de equidad o mecanismos de corrección obligatorios. La ley no exige suspender o modificar el uso de sistemas con impacto dispar comprobado (*disparate impact*), que es el término utilizado en el derecho estadounidense para la discriminación indirecta, sino que se limita a requerir la publicación de auditorías<sup>95</sup>. Esto ha llevado a algunos comentaristas a criticar que el enfoque de la ley está centrado en la transparencia, que presume que con acceso a información pública las empresas adoptarán mejores prácticas, pero sin prever incentivos, reglas de presunción robusta ni sanciones proporcionales al daño causado<sup>96</sup>.

<sup>92</sup> CENTER FOR DEMOCRACY AND TECHNOLOGY, (2022), pp. 10-11.

<sup>93</sup> WRIGHT et al. (2024), pp. 1702-1703.

<sup>94</sup> LAM et al. (2024).

<sup>95</sup> LAM et al. (2024).

<sup>96</sup> GROVES et al. (2024).

Otro aspecto controversial es que la ley focaliza el análisis de sesgo únicamente en variables como raza y sexo, dejando fuera otras categorías protegidas como edad, discapacidad o religión<sup>97</sup>. Además, la responsabilidad recae principalmente en los empleadores que usan las AEDT, y no en los desarrolladores o proveedores de estas herramientas, quienes suelen tener mayor conocimiento y control técnico sobre su funcionamiento<sup>98</sup>. También se ha señalado que el estándar estadístico obligatorio (la regla de los cuatro quintos, utilizada tradicionalmente por la EEOC) podría no ser adecuado en todos los contextos y facilitar la manipulación o cumplimiento superficial de los criterios técnicos<sup>99</sup>. Por último, otro desafío que plantea la LL144NYC en cuanto a su aplicación es la falta de un sistema centralizado y coherente para la recolección de la información sobre el uso de AEDTs y la publicación de los informes de auditoría<sup>100</sup>. La ausencia de un repositorio centralizado obstaculiza la capacidad de los postulantes y trabajadores para acceder a la información necesaria para identificar posibles prácticas discriminatorias, y para luego utilizar esta información como prueba.

### *3.3. Críticas a la implementación y aplicación práctica*

En la fase de aplicación, las críticas se centran en la baja visibilidad pública de las auditorías realizadas, lo que pone en duda el cumplimiento efectivo de la ley. Los auditores han reportado dificultades significativas para acceder a los datos necesarios, tanto de los empleadores como de los proveedores de software, lo cual limita la profundidad y validez de los análisis. También existe confusión sobre la legitimidad de distintos tipos de auditores, en parte debido a la falta de estandarización metodológica, y se ha observado una tendencia a que las empresas encarguen auditorías internas que podrían no cumplir con los principios de independencia exigidos<sup>101</sup>.

Asimismo, los cambios introducidos por el Departamento de Protección al Consumidor y del Trabajador (DCWP) en las métricas estadísticas requeridas han generado retrasos e incertidumbre durante el proceso de adaptación. Por último, se ha criticado que la autoridad de aplicación recae en un órgano con competencia general, y no en la Comisión de Derechos Humanos de la ciudad, lo que debilita las capacidades de fiscalización especializada. Todo esto ha alimentado el riesgo de que la ley sea percibida como un mero ejercicio de cumplimiento formal, sin generar transformaciones sustantivas en la equidad de los procesos de selección algorítmica<sup>102</sup>.

## **4. Conclusiones**

Este trabajo ha propuesto una aproximación crítica a los desafíos probatorios que enfrentan las víctimas de discriminación algorítmica, argumentando que dichos desafíos no solo son técnicos o procedimentales, sino que están enraizados en una concepción del injusto discriminatorio centrado en la víctima. A partir de este enfoque, y de una lectura relacional de la justicia, se mostró cómo los sistemas automatizados de decisión amplifican desigualdades históricas mediante mecanismos difícilmente rastreables o explicables desde el punto de vista jurídico convencional.

Desde el punto de vista científico, el artículo contribuye a una creciente literatura interdisciplinaria que vincula el derecho antidiscriminación, los estudios críticos de tecnología y

<sup>97</sup> En efecto, la regulación de la LL144NYC queda determinada por las categorías protegidas incluidas en el informe del Componente 1 de la EEO. Este es una recopilación de datos anual obligatoria que requiere que todos los empleadores del sector privado con 100 o más empleados, y los contratistas federales con 50 o más empleados que cumplan ciertos criterios, envíen datos demográficos de la fuerza laboral, desagregados por categoría de trabajo, sexo y raza o etnia. EEOC (2023).

<sup>98</sup> GROVES et al. (2024).

<sup>99</sup> Esto ha sido analizado en detalle en CODDOU et al. (2025); y en CODDOU Y PADILLA (2024).

<sup>100</sup> WRIGHT et al. (2024), p. 1709.

<sup>101</sup> WRIGHT et al. (2024), p. 1704.

<sup>102</sup> WRIGHT et al. (2024), p. 1709.

los debates filosóficos sobre la justicia. Particularmente, propone que una concepción relacional del daño discriminatorio permite repensar la carga de la prueba, desplazando la exigencia de demostrar intención o causalidad hacia la necesidad de visibilizar impactos sistémicos y desproporcionados. Desde el punto de vista institucional, se evidencia la necesidad de repensar los marcos regulatorios existentes para enfrentar los riesgos de opacidad, asimetría de información y captura técnica que afectan los procesos automatizados. Las iniciativas como la LL144NYC representan avances importantes, pero insuficientes: su énfasis en la transparencia y la autoevaluación por parte de los empleadores muestra los límites de un enfoque basado exclusivamente en la publicación de auditorías, sin mecanismos efectivos de corrección, sanción ni participación de los afectados. De todos modos, avanzar en cuestiones fundamentales en materia de transparencia, permitirá realizar escrutinios que hasta el momento están ausentes del debate público, como la posibilidad de auditar la validez, confiabilidad y representatividad de los datos de entrenamientos de los PARS o las evaluaciones *ex post* que pueden realizarse una vez que los PARS se implementan en conjuntos de datos que difieren de manera fundamental respecto de los datos de entrenamiento. Esto permitirá cuestionar no sólo la racionalidad científica de los PARS, sino evaluar si acaso estos sistemas son realmente capaces de cumplir lo que prometen cuando ofertan sus servicios en el mercado de los recursos humanos.

A partir del análisis desarrollado en este artículo, surgen diversas preguntas de investigación que pueden orientar futuras agendas de estudio en el campo del derecho antidiscriminación en contextos de automatización. En primer lugar, resulta crucial indagar cómo los marcos jurídicos pueden redistribuir de manera efectiva la carga probatoria cuando existe una radical asimetría de información entre las personas afectadas y quienes controlan los sistemas algorítmicos. Esto exige examinar no sólo reformas procesales, sino también estructuras y actores institucionales capaces de acompañar a las víctimas en la producción de evidencia. En segundo lugar, se vuelve urgente explorar qué tipos de métricas de equidad estadística podrían capturar de forma justa y contextualizada los impactos discriminatorios de los algoritmos, sin reducir el análisis a estándares estadísticos simplificados que pueden ocultar dinámicas estructurales de exclusión. Asimismo, cabe preguntarse qué modelos de auditoría y evaluación de impacto *ex ante* serían adecuados para incorporar no sólo a expertos técnicos, sino también a actores públicos y organizaciones de la sociedad civil que representen el interés colectivo en estas materias. Vinculado a lo anterior, otra línea de investigación relevante dice relación con la posibilidad de diseñar mecanismos institucionales de reparación colectiva en contextos donde los daños no siempre son identificables de manera individual ni atribuibles a una única decisión. Finalmente, se requiere examinar los límites legítimos que deben establecerse entre la protección de la propiedad intelectual y el derecho de las personas a una explicación significativa sobre las decisiones automatizadas que las afectan. Estas preguntas no sólo son relevantes desde un punto de vista teórico, sino que constituyen desafíos concretos para el desarrollo de sistemas jurídicos capaces de responder a las nuevas formas de desigualdad en la era digital.

#### BIBLIOGRAFÍA CITADA

ACCATINO, DANIELA (2021): “La exigencia de motivación de las decisiones probatorias como garantía”, en: A.A.V.V, Manual sobre derechos humanos y prueba en el proceso penal (Ciudad de México, Suprema Corte de Justicia de la Nación) pp. 267-307.

ADAMS-PRASSL, JEREMIAH; BINNS, REUBEN Y KELLY-LYTH, ANNE (2023): “Directly discriminatory algorithms”, en: The Modern Law Review (Vol. 86, N° 1), pp. 144-175.

ALSTON, PHILLIP (2018): “Statement on visit to the United Kingdom by Philip Alston, United Nations Special Rapporteur on Extreme Poverty and Human Rights”. Disponible en:

[https://www.ohchr.org/sites/default/files/Documents/Issues/Poverty/EOM\\_GB\\_16Nov2018.pdf](https://www.ohchr.org/sites/default/files/Documents/Issues/Poverty/EOM_GB_16Nov2018.pdf) [visitado el 2 de noviembre de 2023].

ANDERSON, ELIZABETH (1999): “What is the Point of Equality?”, en: *Ethics* (Vol. 109, N° 2), pp. 287-337.

ARTICLE 29 DATA PROTECTION WORKING PARTY (2018): “Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679”. Disponible en: <https://ec.europa.eu/newsroom/article29/items/612053/en> [visitado el 4 de octubre de 2023].

AJUNWA, IFEOMA (2023): *The quantified worker: Law and technology in the modern workplace* (Nueva York, Cambridge University Press).

ANDREWS, CAITLIN (2024): “How the US is handling AI-driven hiring practices”. Disponible en: <https://iapp.org/news/a/how-the-us-is-handling-ai-driven-hiring-practices> [visitado el 27 de mayo de 2025].

ANDREWS, CAITLIN (2025): “European Commission withdraws AI Liability Directive from consideration”. Disponible en: <https://iapp.org/news/a/european-commission-withdraws-ai-liability-directive-from-consideration> [visitado el 17 de abril de 2025].

AZUAJE PIRELA, MICHELLE Y FINOL, DANIEL (2023): “Aproximaciones a la noción de inteligencia artificial y otros conceptos vinculados con ella”, en: Michelle Azuaje-Pirela (Ed.), *Introducción a la ética y el derecho de la inteligencia artificial* (Madrid, La Ley), pp. 17-34.

BAMFORTH, NICHOLAS; MALIK, MALEIHA Y O'CINNEIDE, COLM (2008): *Discrimination Law: Theory and Context (Text and Materials)* (Londres, Thomson: Sweet & Maxwell).

BAROCAS, SOLON Y SELBST, ANDREW (2016): “Big data’s disparate impact”, en: *California Law Review* (Vol. 104, N° 3), pp. 671-732.

BOGEN, MIRANDA (2019): “All the ways hiring algorithms can introduce bias», *Harvard Business Review*. Disponible en: <https://bit.ly/3UHmpoq> [visitado el 10 de abril de 2025].

CALVI, ALESSANDRA (2023): “Exploring the Synergies between Non-Discrimination and Data Protection: What Role for EU Data Protection Law to Address Intersectional Discrimination?”, en: *European Journal of Law and Technology* (Vol. 14, N° 2). Disponible en: <https://ejlt.org/index.php/ejlt/article/view/912> [visitado el 3 de marzo de 2025].

CENTER FOR DEMOCRACY AND TECHNOLOGY (2022): “Algorithm Driven Hiring Tools: Innovative Recruitment or Expedited Disability Discrimination?”. Disponible en: <https://cdt.org/wp-content/uploads/2020/12/Full-Text-Algorithm-driven-Hiring-Tools-Innovative-Recruitment-or-Expedited-Disability-Discrimination.pdf> [visitado el 19 de marzo de 2025].

CODDOU, ALBERTO (2018): “A Transformative Approach to Anti-Discrimination Law in Latin America. Doctoral dissertation, UCL”, en: University College London. Disponible en: [https://discovery.ucl.ac.uk/id/eprint/10064268/1/Coddou%20McManus\\_\\_PhD%20\(after%20corrections%20December%202018\)%20E%20Thesis.pdf](https://discovery.ucl.ac.uk/id/eprint/10064268/1/Coddou%20McManus__PhD%20(after%20corrections%20December%202018)%20E%20Thesis.pdf) [visitado el 3 de noviembre de 2023].

CODDOU, ALBERTO Y PADILLA, ROBERTO (2024): “Discriminación algorítmica en los procesos automatizados de reclutamiento y selección de personal”, en: *Revista chilena de derecho y tecnología* (N° 13), pp. 1-34.

COHEN, G.A. (1989): “On the currency of egalitarian justice”, en: *Ethics* (Vol. 99, N° 4), pp. 906-944.

COLLINS, HUGH Y KHAITAN, TARUNABH (2018): *Foundations of indirect discrimination law* (Londres, Bloomsbury Publishing).

COLOMA, RODRIGO. (2017): “Conceptos y razonamientos probatorios”, en: *Revista de derecho* (Valdivia) (Año 30, N° 2), pp. 31-56.

COMISIÓN EUROPEA (2022): “Preguntas y respuestas: Directiva sobre responsabilidad civil en materia de IA”. Disponible en: [https://ec.europa.eu/commission/presscorner/detail/es/qanda\\_22\\_5793](https://ec.europa.eu/commission/presscorner/detail/es/qanda_22_5793) [visitado el 5 de abril de 2025].

CONTRERAS, PABLO Y VIOLLIER, PABLO (en prensa): “Derecho a una explicación y regulación de las decisiones automatizadas: contrapunto entre el Reglamento General de Protección de Datos y la reforma chilena en protección de datos personales”.

COUTURE, EDUARDO (1988): Fundamentos del derecho procesal civil, 3<sup>a</sup> edición (Buenos Aires, ediciones Depalma).

EDWARDS, LILIAN Y VEALE, MICHAEL (2017): “Slave to the algorithm? Why a right to an explanation is probably not the remedy you are looking for”, en: Duke Law & Technology Review (Vol 16, N° 18), pp. 18-84.

EEOC (2023): “EEO-1 Component 1 Data Collection”. Disponible en: <http://www.eeodata.org/eeo1> [visitado el 15 de mayo de 2025].

FJELD, JESSICA; ACHTEN, NELE; HILLIGOS, HANNAH; NAGY, ADAM Y SRIKU-MAR, MADHULIKA (2020): “Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI”, en: Berkman Klein Center for Internet & Society. Disponible en: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3518482](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482) [visitado el 12 de diciembre de 2023].

FUCHS, LINDSEY (2023): “Hired by Machine: Can a New York City Law Enforce Algorithmic Fairness in Hiring Practices?”, en: Fordham Journal of Corporate and Financial Law (N° 28), pp. 185-222.

GINÈS I FABRELLAS, ANA (2021): “Sesgos discriminatorios en la automatización de decisiones en el ámbito laboral: Evidencias de la práctica”, en: Rivas, Pilar (Ed.), Discriminación algorítmica en el ámbito laboral: Perspectiva de género e intervención (Madrid, Thomson Reuters-Aranzadi), pp. 295-331.

GOLDBERG, SUZANNE (2010): “Discrimination by comparison”, en: Yale Law Journal (N° 120), pp. 728-812.

GROVES, LARA; METCALF JACOB; KENNEDY, ALAYNA; VECCHIONE, BRIANA Y STRAIT, ANDREW (2024): “Auditing Work: Exploring the New York City algorithmic bias audit regime”, en: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (Association for Computing Machinery, New York, NY, USA), pp. 1107-1120.

HELLMAN, DEBORAH (2008): When is discrimination wrong? (Cambridge, Harvard University Press).

HELLMAN, DEBORAH (2020): “Measuring algorithmic fairness”, en: Virginia Law Review (Vol. 106, N° 4), pp. 811-866.

HELLMAN, DEBORAH Y MOREAU, SOPHIA (2013): Philosophical foundations of discrimination law (Oxford, OUP).

KAMINSKI, MARGOT (2021): “The right to explanation, explained”, en: Research handbook on information law and governance (Londres, Edward Elgar Publishing), pp. 278-299.

KELLY-LYTH, AISLINN (2021): “Challenging biased hiring algorithms”, en: Oxford Journal of Legal Studies (Vol. 41, N° 4), pp. 899-928.

KELLY-LYTH, AISLINN (2023): “Algorithmic discrimination at work”, en: European Labour Law Journal (Vol. 14, N° 2), pp. 152-171.

KIM, PAULINE (2017): “Auditing algorithms for discrimination”, en: University of Pennsylvania Law Review (N° 166), pp. 189-20.

KIM, PAULINE (2022): “Race-aware algorithms: Fairness, nondiscrimination and affirmative action”, en: California Law Review (N° 110), pp. 1539-1596.

KITCHIN, ROB (2017): "Thinking critically about and researching algorithms", en: *Information, Communication & Society* (Vol. 20, N° 1), pp. 14-29.

KUPFER, CORDULA; PRASSL, RITA; FLEIß, JURGEN; MALIN, CHRISTINE; THALMANN, STEFFAN Y KUBICEK, BETTINA (2023): "Check the box! How to deal with automation bias in AI-based personnel selection", en: *Frontiers in Psychology* (N° 14), pp. 1-26.

LAM, KHOA; LANGE, BENJAMIN; BLILI-HAMELIN, BORHANE; DAVIDOVIC, JOVANA; BROWN, SHEA Y HASAN, ALI (2024): "A Framework for Assurance Audits of Algorithmic Systems", en: *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. Disponible en: <https://doi.org/10.1145/3630106.3658957> [visitado el 19 de marzo de 2025].

LIPPERT-RASMUSSEN, KASPER (2013): *Born free and equal?: A philosophical inquiry into the nature of discrimination* (Oxford, OUP).

MALGIERI, GIANCARLO (2019): "Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations", en: *Computer law & security review* (Vol. 35, N° 5), pp. 1-26.

MOREAU, SOPHIA (2010): "What is discrimination?", en: *Philosophy & Public Affairs* (N° 38), pp. 143-179.

MOREAU, SOPHIA (2020): "Equality and Discrimination", en: Tasioulas, John (Ed.), *The Cambridge Companion to the Philosophy of Law* (Cambridge, Cambridge University Press), pp. 171-190.

MOREAU, SOPHIA (2024): "Beyond Discrimination Law: Realizing Equality Through Other Laws, Such as Tort Law", en: *American Journal of Law and Equality* (N° 4), pp. 427-461.

MUÑOZ, FERNANDO (2015): "Estándares conceptuales y cargas procesales en el litigio antidiscriminación. Análisis crítico de la jurisprudencia sobre Ley Zamudio entre 2012 y 2015", en: *Revista de derecho* (Valdivia) (Año 28, N° 2), pp. 145-167.

NARAYANAN, ARVIND Y KAPOOR, SAYASH (2024): *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference* (Princeton, NJ, Princeton University Press).

NAUGHTON, JON (2022): "Computers can write their own code. So are programmers now obsolete?". Disponible en: <https://www.theguardian.com/commentisfree/2022/feb/12/computers-can-write-their-own-code-so-are-programmers-now-obsolete> [visitado el 17 de abril de 2025].

NOBLE, SAFIYA (2018): *Algorithms of oppression: How search engines reinforce racism* (New York, NYU Press).

NEVEN, JEAN-FRANÇOIS; RINGELHEIM, JULIE Y VAN DER PLANCKE, VÉRONIQUE (2022): "Les tests de situation et la méthode du client-mystère comme outils de détection et de preuve des discriminations", en: CRIDHO Working Paper (N° 5). Disponible en: [https://sites.uclouvain.be/cridho/documents/Working.Papers/CRIDHO-WP-2022-5\\_Neven-Ringelheim-vanderPlancke\\_Tests%20situation.pdf](https://sites.uclouvain.be/cridho/documents/Working.Papers/CRIDHO-WP-2022-5_Neven-Ringelheim-vanderPlancke_Tests%20situation.pdf) [visitado el 24 de enero de 2024].

NORDELL, JESSICA (2021): *The end of bias: A beginning: The science and practice of overcoming unconscious bias* (Nueva York, Metropolitan Books).

NYC CONSUMER AND WORKER PROTECTION (2023): "Automated Employment Decision Tools (AEDT)". Disponible en: <https://www.nyc.gov/site/dca/about/automated-employment-decision-tools.page> [visitado el 19 de marzo de 2025].

ORGANIZACIÓN INTERNACIONAL DEL TRABAJO (2021): "Perspectivas Sociales y del Empleo en el Mundo 2021: El papel de las plataformas digitales en la transformación del mundo del trabajo". Disponible en:

[https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@dgreports/@dcomm/@publ/documents/publication/wcms\\_823119.pdf](https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@dgreports/@dcomm/@publ/documents/publication/wcms_823119.pdf) [visitado el 23 de diciembre de 2023].

SERRANO FALCÓN, CAROLINA (2021): "Proceso de contratación, política de empleo y uso de algoritmos", en: Rivas, Pilar (Ed.), *Discriminación algorítmica en el ámbito laboral: Perspectiva de género e intervención* (Madrid, Thomson Reuters-Aranzadi), pp. 179-209.

STETZ, THOMAS (2022): *Test Bias in Employment Selection Testing: A Visual Introduction* (Cham, Springer International Publishing).

VILJOEN, SALOME (2021): "A relational theory of data governance", en: *The Yale Law Journal* (Nº 131), pp. 573-654.

VIOLIER, PABLO Y FISCHER, EMMA (2023): "La intervención humana como resguardo ante la toma automatizada de decisiones: implicancias éticas y jurídicas", en: Azuaje-Pirela, Michelle (Ed.), *Introducción a la ética y el derecho de la inteligencia artificial* (Santiago, La Ley), pp. 151-159.

WACHTER, SANDRA (2022): "The theory of artificial immutability: Protecting algorithmic groups under anti-discrimination law", en: *Tulane Law Review* (Vol. 97, Nº 2), pp. 1-50.

WACHTER, SANDRA; MITTELSTADT, BRENT Y RUSSELL, CHRIS (2021): "Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI", en: *Computer Law & Security Review* (Nº 41), pp. 1-72.

WRIGHT, LUCAS; MIKA MUENSTER, ROXANA; VECCHIONE, BRIANA,; TIANYAO QU, PIKA (SENHUANG) CAI Y SMITH, ALAN (2024): "Null Compliance: NYC Local Law 144 and the challenges of algorithm accountability", en: *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Disponible: <https://doi.org/10.1145/3630106.3658998> [visitado el 19 de febrero de 2025].

WORLD ECONOMIC FORUM (2022): *Global Risks Reports, 17<sup>th</sup> edition* (Geneva, World Economic Forum).

XENIDIS, RAPHAEL (2020): "Tuning EU equality law to algorithmic discrimination: Three pathways to resilience", en: *Maastricht Journal of European and Comparative Law* (Vol. 27, Nº 6), pp. 736-758.

XENIDIS, RAPHAEL (2023): "Beyond bias: algorithmic machines, discrimination law and the analogy trap", en: *Transnational Legal Theory* (Vol. 14, Nº 4), pp. 378-412.

ZUIDERVEEN, FREDERIK (2020): "Strengthening legal protection against discrimination by algorithms and artificial intelligence", en: *The International Journal of Human Rights* (Vol. 24, Nº 10), pp. 1572-1593.

#### JURISPRUDENCIA CITADA

Tribunal de Justicia de la UE, *Tadao Maruko contra Versorgungsanstalt der deutschen Bühnen*, C-267/06, 1 de Abril de 2008.

Estados Unidos, Supreme Court, *Wal-Mart v. Dukes*, 10-277 U.S. (2011).

Tribunal de Justicia de la UE, *CHEZ Razpredelenie Bulgaria AD y Komisia za zashtita ot diskriminatsia*, C-83/14, 16 de Julio de 2015.

Reino Unido, Corte Suprema, *R (on the application of Coll) v Secretary of State for Justice* (2017) UKSC 40.

Estados Unidos, District Court of Columbia, *Sandvig v. Barr* (2020).

#### NORMAS JURÍDICAS CITADAS

Reglamento General de Protección de Datos o Reglamento 2016/679. Diario Oficial de la Unión Europea, 4 de mayo de 2016.

Code du Travail de Francia. 22 de septiembre de 2017.

Equality Act 2010 del Reino Unido. Consentimiento real, 8 de abril de 2010.

Ley N° 18.620, Código del Trabajo. Diario oficial, 1931.

Public Law N° 99-474 Computer Fraud and Abuse Act, Title 18, United States Code, Section 1030. Office of the Federal Register, 8 de octubre de 1986.

2021/144 NYC Local Law of 2021. Publicada el 1 de enero de 2023.

Public Law 107-204 Sarbanes-Oxley Act of 2002. Publicada el 30 de julio de 2002.