

DOCTRINA

La discriminación en una sociedad automatizada: Contribuciones desde América Latina

Discrimination in an automated society: Contributions from Latin America

Catherine Muñoz Gutiérrez 

Abogada independiente, Chile

RESUMEN ¿Las actuales legislaciones latinoamericanas protegen adecuadamente a las personas frente a discriminaciones por el uso de sistemas de inteligencia artificial de toma de decisiones automatizadas?, ¿cómo se debe configurar un marco legal efectivo sobre esta materia? A partir de estas preguntas, el presente artículo analiza este tipo de discriminación desde el punto de vista técnico y regulatorio, desarrolla una teoría conceptual sobre las particularidades de una discriminación algorítmica situada y contextualizada en América Latina, propone la incorporación de un principio relacional propio de la región y ofrece un nuevo enfoque interpretativo del principio de igualdad y del concepto de discriminación desde una perspectiva colectiva, con el objeto de superar dificultades normativas evidenciadas en el derecho comparado.

PALABRAS CLAVE Principio de igualdad, discriminación, inteligencia artificial, América Latina, toma de decisiones automatizadas.

ABSTRACT Do current Latin American laws adequately protect people against discrimination due to the use of artificial intelligence systems for automated decision-making? How should an effective legal framework on this matter be configured? Based on these questions, this article analyzes this form of discrimination from a technical and regulatory point of view, develops a conceptual theory on the particularities of algorithmic discrimination situated and contextualized in Latin America, proposes the incorporation of a relational principle that is distinctive to the region and offers a new interpretative approach to the Principle of Equality and the discrimination concept, from a collective perspective, with the aim of overcoming normative difficulties evidenced in comparative law.

KEYWORDS Principle of equality, discrimination, artificial intelligence, Latin America, automated decision-making.

Introducción

La injusticia en cualquier lugar es una amenaza para la justicia en todas partes. Estamos atrapados en una red ineludible de mutualidad, atados en una única prenda del destino. Lo que afecta a uno directamente, afecta a todos indirectamente.

MARTIN LUTHER KING, JR.¹

A mediados de marzo del año 2020, dentro de las medidas adoptadas para enfrentar la pandemia del Covid-19, el Primer Ministro británico anunció el cierre de las escuelas y la cancelación de los exámenes de admisión universitaria, GCSE y A-level. Para su reemplazo la Oficina de Calificaciones (Ofqual) desarrolló un sistema para determinar un puntaje de ingreso universitario para cada estudiante de forma predictiva y automatizada, utilizando inteligencia artificial. El resultado estaba supeditado principalmente a dos variables: 1) datos estadísticos relacionados con el desempeño histórico en cada establecimiento educacional, esto es, el desempeño de estudiantes de años anteriores y no directamente el desempeño de aquellos jóvenes sometidos a este proceso; y 2) un puntaje estimado dado por cada profesor en base a un sistema de calificación previamente estandarizado.²

Las referidas variables estaban ajustadas de manera que los registros estadísticos de desempeño por establecimiento tenían mayor peso que el puntaje estimado de profesores con el objeto de disminuir cualquier inflación de calificaciones. Con la publicación de los resultados quedó en evidencia que los estudiantes de sectores pobres y vulnerables fueron drásticamente perjudicados en su puntaje, con un promedio de un 40% por debajo de sus calificaciones regulares, mientras que aquellos alumnos de clases acomodadas se beneficiaron llegando a una proporción de notas A* y A histórica que dobló la tasa de las escuelas públicas. El sistema asignó mayor valor a evaluaciones emitidas por profesores pertenecientes a escuelas pequeñas ya que, según informó el Gobierno, no existía un modelo estadístico seguro para aplicar la variable de registros de desempeño histórico respecto de grupos reducidos de estudiantes. Este tipo de escuelas reducidas están directamente relacionadas con establecimientos privados exclusivos para grupos de altos ingresos. En contraste, un grupo históricamente en desventaja, en este caso, jóvenes pertenecientes a familias de bajos ingresos, fueron desproporcionadamente perjudicados, reproduciendo y amplificando una desigualdad social y evidenciando una discriminación, descartando con ello su real capacidad y desempeño, los cuales fueron reemplazados por

1. Véase Martin Luther King Jr., «Letter from Birmingham jail, April 16, 1963». University of Pennsylvania African American Studies Center, disponible en <https://bit.ly/35ECFw6>.

2. Véase Paul Johnson, «Method Used to Assign A-Level Grades favoured Public Schools», *The Times*, 17 de agosto de 2020, disponible en <https://bit.ly/3hXYDSa>.

un valor dado en base a un modelo estadístico de registros de otros estudiantes de años anteriores.³

Este no es un caso aislado, hay un patrón detectado en el uso de sistemas de inteligencia artificial (IA) automatizados —que asisten o sustituyen procesos de toma de decisiones que, de otro modo, serían realizados por humanos—⁴ cuando se emplean en materias de relevancia social. En dichos casos, grupos históricamente excluidos y oprimidos, tales como personas afroamericanas, latinas, pueblos originarios, comunidades LGBTIQ+, minorías religiosas, personas de escasos recursos, entre otras, son comúnmente afectados de forma desproporcionada, reproduciendo y perpetuando injusticias sociales (Costanza-Chock, 2020; D'Ignazio y Klein, 2020; Zuboff, 2019; Benjamin, 2020).

Sistemas predictivos de abuso infantil que asocian injustamente pobreza con abusos (Eubanks, 2018: 127), sistemas predictivos de delitos que asocian delincuencia a racismo policial histórico (Selbst, 2017), sistemas predictivos que detectan erróneamente fraudes de beneficios sociales a personas honestas de escasos recursos (Gilman, 2020), la determinante y peligrosa discriminación en sistemas de reconocimiento facial (Buolamwini y Gebru, 2018), los generalizados sistemas automatizados de contratación de personal (Sánchez-Monedero, Dencik y Edwards, 2020), son solo algunos ejemplos.

Evaluar y fortalecer los actuales marcos legales para la regulación de este tipo de sistemas de IA se ha vuelto prioritario (Pasquale, 2017; Wachter, Mittelstadt y Russell, 2020; Barocas y Selbst, 2016) considerando el aumento exponencial del uso de este tipo de soluciones en el sector privado (Creel y Hellman, 2021), así como el creciente interés de incorporar dentro de políticas públicas sistemas de toma de decisiones automatizadas para la administración del Estado⁵ en materias como prevención del delito, seguridad social, justicia, salud, educación y subsidios públicos (Brown y otros, 2019).

Hasta ahora las distintas propuestas de regulación sobre esta materia han sido desarrolladas principalmente en Europa (Goodman y Flaxman, 2017), Canadá (McKelvey y MacDonald, 2019) y algunos estados norteamericanos.⁶

3. Véase Lauri Clarke, «How The A-Level Results Algorithm Was Fatally Flawed», *NS Tech*, 14 de Agosto de 2020, disponible en <https://bit.ly/3fA7KHb>.

4. Véase «Algorithmic Accountability Policy Toolkit», *AI Now*, octubre de 2018, disponible en <https://bit.ly/2UmSu8k>.

5. La Estrategia de Transformación Digital del Gobierno de Chile del año 2018 señala como uno de sus objetivos el potenciar la optimización de políticas públicas y la automatización de procesos mediante el uso de tecnologías emergentes tales como la inteligencia artificial. Véase «Estrategia De Transformación Digital 2018-2020», División de Gobierno Digital del Ministerio Secretaría General de la Presidencia de Chile, 2018, disponible en <https://bit.ly/3cYFFaO>.

6. El estado de Nueva York cuenta con varios ejemplos de propuestas legislativas en materia de sistemas de decisiones automatizadas por IA. Un ejemplo de estas iniciativas se puede encontrar en el

Es importante destacar que este particular tema regulatorio posee una notoria complejidad y, contrario a la intuición, esta no se relaciona con la complejidad tecnológica, sino por la propia materia legal abordada. En efecto, las normas sobre igualdad y no discriminación se han construido bajo complejas e históricas tensiones políticas y sociales, existiendo, al día de hoy, diversas teorías e interpretaciones (Khaitan, 2015). Los conceptos de igualdad y discriminación tienen distintos contenidos y alcances en cada sistema jurídico, (Hagerty y Rubinov, 2019) poseyendo, además, particularidades propias en cada país, siendo en definitiva conceptos elaborados en virtud de su contexto social (Hellman y Moreau, 2013). De la misma manera, el concepto antítesis de discriminación del norte global, *fairness*, posee distintas interpretaciones dependiendo si se usa en el área computacional, social o legal (Xiang y Raji, 2019a), con distinto alcance según el contexto social de cada país (Sambasivan y otros, 2021) y no cuenta con un equivalente exacto en Latinoamérica, siendo interpretado como imparcialidad, equidad o justicia (Rawls, 1990).

Pues bien, teniendo presente lo anterior, este artículo realiza las siguientes contribuciones: 1) ofrece un análisis sobre las particularidades técnicas y legales de una problemática poco abordada en Latinoamérica; 2) propone una teoría conceptual sobre la discriminación algorítmica situada en el contexto latinoamericano, identificando las particularidades de grupos protegidos, prácticas discriminatorias propias de la región y desafíos asociados; 3) propone la adopción de un principio relacional de corte filosófico, presente transversalmente en todos los pueblos originarios latinoamericanos, cuya visión y contenido permite afrontar apropiadamente los desafíos que presenta la IA, y 4) propone un nuevo enfoque interpretativo de los conceptos de igualdad y discriminación con un contenido colectivo y ético moral, en línea con el referido principio relacional. Estas interpretaciones superan las dificultades que presentan las actuales normas de la materia en comento, teniendo por finalidad, además, no solo condenar o mitigar una discriminación, sino superar el statu quo y dismantelar las estructuras de injusticia social que las originan. Lo propuesto sirve de base para una evaluación, reinterpretación y fortalecimiento legal en esta materia.

Este análisis sigue la siguiente estructura: entendiendo que en materia de IA existen múltiples interpretaciones y puntos de vista, se ha considerado necesario incorporar en el primer apartado elementos básicos relacionados con esta tecnología, describiendo el actual nivel de desarrollo tecnológico, aspectos elementales del centro de este desarrollo, el aprendizaje automático o *machine learning* (ML) y particularidades

proyecto de ley Int 1806-2019, «A Local Law to amend the administrative code of the city of New York, in relation to reporting on automated decision systems used by city agencies, Int. 1806 (2019)», actualmente en trámite legislativo, disponible en <https://on.nyc.gov/35HwjMm>.

sobre los sistemas de ML utilizados para la toma de decisiones automatizadas. La segunda parte de este primer apartado explora la actual tendencia a la hora de conceptualizar la IA para propósitos regulatorios, destacando el aspecto sociotécnico de este concepto. El segundo apartado analiza cómo los sistemas de IA de toma de decisiones automatizadas pueden generar discriminación. Este acápite se desarrolla de forma complementaria desde el punto de vista técnico y desde la relación que poseen estas herramientas de IA con el poder presente en las sociedades,⁷ tomando para ello perspectivas teóricas feministas y decoloniales.

El tercer apartado hace un análisis exploratorio de las actuales normas legales sobre la materia de estudio en el derecho comparado, en particular, normas europeas y norteamericanas, detectando y describiendo debilidades y desafíos legales. El cuarto apartado desarrolla una teoría conceptual sobre la discriminación algorítmica situada y contextualizada en América Latina, caracterizando particularidades sobre grupos protegidos, prácticas discriminatorias particulares y desafíos frente a este nuevo tipo de discriminación. El quinto apartado propone la adopción de un principio relacional filosófico cuya visión y contenido dan herramientas y perspectivas para una relación constructiva con este tipo de tecnología y da cuenta de una nueva interpretación de los conceptos de igualdad y discriminación. Finalmente, el sexto apartado pone en práctica la construcción descriptiva establecida en el punto anterior, tomándola como parámetro para el análisis y evaluación empírica de un caso de estudio: la normativa vigente en Chile.

Este trabajo se limita al estudio regulatorio exclusivamente de sistemas de IA de toma de decisiones automatizadas que producen consecuencias discriminatorias, desde las normas de no discriminación.⁸ Para efectos prácticos e ilustrativos este análisis se ha centrado en aspectos relacionados con el racismo, pudiendo ser sus conclusiones igualmente aplicadas a otras características de grupos protegidos. En vista de lo anterior, y en pos de preservar la legibilidad de este trabajo, términos como sistemas de IA, discriminación algorítmica, algoritmos o modelos se circunscriben y refieren exclusivamente a este tipo de sistemas y temática particular.

7. El poder se refiere al control que es ejercido entre individuos o grupos respecto de recursos intelectuales, humanos y financieros, siendo dinámico y relacional y desplegado en el ámbito social, económico y político. Para más información, véase VeneKlasen y otros (2002).

8. Al respecto, se hace presente que la amplitud de un marco legal para sistemas de inteligencia artificial es multisectorial, extenso vinculada a normas internacionales sobre derechos humanos y circunscrito a diversas áreas del derecho, las cuales, además de las normas sujetas al presente análisis, cubren materias como protección al consumidor, responsabilidad civil, libre competencia, defensa e inteligencia, normas administrativas, entre otras.

Inteligencia artificial desde la perspectiva regulatoria: Aspectos básicos

Contexto actual de desarrollo de la IA

La representación tecnológica está estrechamente conectada con la visión que tenemos sobre nuestro futuro como sociedad. Respecto de la IA, existe una narrativa exagerada sobre estas expectativas. Constantemente se informa sobre la capacidad de la IA para mejorar nuestro futuro, predecir nuestro comportamiento y solucionar graves problemas sociales. Se habla de sistemas inteligentes autónomos, que aprenden y pueden reemplazar completamente funciones cognitivas humanas.

Lo anterior guarda relación con la denominada inteligencia artificial general (AGI, por sus siglas en inglés), que corresponde a una hipótesis sobre el desarrollo y estudio de inteligencias artificiales a nivel humano y con capacidad de generalización (Hagerty y Rubinov, 2019). La AGI, que es la que vemos frecuentemente en la ciencia ficción, no ha sido desarrollada y dista mucho de la tecnología actual, incluso más, se cuestiona la posibilidad de alcanzarla en un corto plazo o derechamente que sea materialmente posible su creación (Fjelland, 2020). Los llamados sistemas autónomos hasta ahora no son tales, existen al respecto importantes limitaciones, además de incluir en todo su proceso significativo trabajo humano (Broussard, 2018; Tubaro y Casilli, 2019).⁹

Lo cierto es que, teniendo presente el desarrollo actual de esta tecnología, se habla de IA estrecha, queriendo con ello describir que corresponden a sistemas que resuelven tareas limitadas y específicas, por ejemplo, reconocer objetos, jugar ajedrez, detectar *spams* o calificar riesgos (Broussard, 2018). Estos sistemas solo sirven para la tarea para la cual han sido programados y no responden a otro escenario, aunque sea solo ligeramente diferente (Marcus, 2020).

En consecuencia, la IA entendida como una rama de las ciencias computacionales¹⁰ (Schuett, 2019) se denomina IA estrecha, en vista a su nivel de desarrollo, la cual se ha centrado principalmente en el aprendizaje automatizado o *machine learning*¹¹

9. Usualmente se utilizan de forma errónea y equivalente los términos automatización y autonomía, siendo diferentes. La automatización ya existe hace bastantes años, el ejemplo clásico es el piloto automático que poseen los aviones; en definitiva, alude a un sistema que realiza tareas definidas para producir resultados determinados, basados en reglas o algoritmos, que no pueden actuar en otros escenarios y sobre los cuales los humanos tienen un papel de supervisor. Por su parte, la autonomía que se relaciona con IA y que no se ha desarrollado en la actualidad guarda relación con sistemas que realizan tareas sin intervención humana, con características cognitivas de adaptación, autoejecución y pudiendo actuar con éxito en eventualidades y otros contextos. Al respecto véase Kaber (2018) y Broussard (2018).

10. La IA como una rama de las ciencias computacionales se divide, a su vez, en múltiples otras ramas, siendo una de ellas el *machine learning*. Para más detalles sobre las distintas ramas que comprende la IA ver Gupta y Nagpal (2020).

11. El desarrollo del ML no es algo nuevo, ha estado presente desde hace más de sesenta años. Al

y , dentro de este, el subgrupo denominado aprendizaje profundo o *deep learning*¹² (Broussard, 2018; Page, Bain y Mukhlis, 2018). Al respecto, Lerh y Ohm definen el ML en los siguientes términos:

Un proceso automatizado de descubrimiento de correlaciones (a veces alternativamente referidos como relaciones o patrones) entre variables en un conjunto de datos, a menudo para hacer predicciones o estimaciones de algún resultado (Lehr y Ohm, 2017).

El funcionamiento básico de un ML se detalla en el siguiente ejemplo de un aprendizaje supervisado:¹³ un algoritmo¹⁴ es programado para obtener un modelo optimizado (una función abstracta matemática) para dar respuesta a una tarea determinada a partir de la identificación de patrones de un conjunto de datos previamente etiquetados, llamados *training data*¹⁵ o datos de entrenamiento. El algoritmo estima las tasas de error de cada modelo aplicando otro conjunto externo de datos, los *validation set* o datos de validación, si el modelo tiene un rendimiento satisfactorio, termina el entrenamiento. Si el rendimiento no produce suficientes resultados, se modifica el modelo y se repita el proceso. El algoritmo selecciona el modelo que mejor describe las características y tendencias a partir de los conjuntos de datos de entrenamiento y validación. Finalmente, mediante un tercer conjunto de datos llamados *test set* o datos de pruebas, se calcula el error de predicción generalizado, lo que queda a disposición del usuario final para una evaluación adecuada del nivel predictivo o análisis del modelo.¹⁶

Por su parte, *deep learning*, en términos simples, corresponde a un tipo de técnica de modelamiento de datos mediante múltiples capas de redes neuronales que corresponden a unidades matemáticas enlazadas. El entrenamiento tiene por objeto ajustar el peso de estos enlaces en base a un conjunto de datos de entrada, para luego funcionar de forma predictiva.¹⁷

Como se ha señalado, este artículo analiza exclusivamente un tipo de ML, los sistemas de tomas de decisiones automatizadas, denominado Automated Decision-

respecto véase Catherine Muñoz, «The Remarkable Evolution Of Artificial Intelligence», *Medium*, 5 de mayo de 2020, disponible en <https://bit.ly/3uxXeo6>.

12. En la actualidad todos los grandes avances en materia de IA radican en el ML y en particular en el *deep learning*. Para más información sobre *deep learning* véase Goodfellow, Bengio y Courville (2016).

13. Sobre distintos tipos de entrenamiento de ML véase Alpaydin (2004).

14. Un algoritmo se define como un conjunto de instrucciones o reglas matemáticas que, especialmente si se dan a una computadora, ayudarán a calcular una respuesta a un problema. Véase <https://dictionary.cambridge.org/dictionary/english/algorithm>.

15. El término *dataset* corresponde al conjunto de datos que generalmente se divide en *training data* y *validation set*. Más detalles en Alpaydin (2004).

16. Véase Pedregosa y otros (2011).

17. Véase «Practical Deep Learning for Coders, V3», *Fast.Ai*, 2019, disponible en <https://bit.ly/3coThBW>.

Making o Automated Decision Systems (ADM Systems, por sus siglas en inglés) y que agrupa dos tipos de sistemas: 1) aquellos que utilizan modelos algorítmicos para el apoyo de toma de decisiones, que finalmente son definidas por humanos; y 2) sistemas cuyas decisiones son completamente automatizadas (Orwat, 2020). Ambas formas poseen los mismos riesgos en relación con la afectación de grupos protegidos, ya que respecto del caso «semiautomatizado», existe una tendencia comprobada que las personas confían más en un juicio de un algoritmo que en el propio cuando estos no son coincidentes (Eubanks, 2018).

Es importante destacar que los modelos implementados en el mundo real pueden ser estáticos o, la regla general, desplegarse de forma continua. Por ejemplo, los sistemas de recomendación y gran parte de los sistemas de toma de decisiones automatizadas funcionan en tiempo real, por lo que requieren una infraestructura adicional. Asimismo, los ML que funcionan a gran escala, y en particular los sistemas de toma de decisiones, también pueden convertirse en sistemas de aprendizaje dinámico u *online*, es decir, sistemas en los que los modelos se vuelven a entrenar de forma regular y automática al recopilar nuevos datos (Lehr y Ohm, 2017).

Concepto de IA para propósitos regulatorios

Las escasas propuestas legislativas sobre IA desarrolladas hasta ahora se conceptualizan, principalmente, a partir de una regulación basada en riesgos (Giuffrida, 2019; Schneeberger, Stöger y Holzinger, 2020) the European Commission published a White Paper on Artificial Intelligence (AI. En términos generales, los marcos regulatorios basados en riesgos tienen dos presupuestos básicos: 1) las normas vinculantes deben hacer referencia concreta a los objetivos a regular y a sus riesgos asociados, y 2) se debe desarrollar un sistema para evaluar dichos riesgos y puntuarlos cuantitativa y cualitativamente. La puntuación permite detectar aquellos sistemas de alto riesgo y regularlos con la finalidad de reducir y minimizar el impacto negativo sobre las personas (Baldwin, Cave y Lodge, 2010).

Sin perjuicio de lo anterior, últimamente se han desarrollado propuestas de marcos regulatorios para sistemas de IA basados en normas vinculantes internacionales sobre derechos humanos (Yeung, Howes y Pogrebna, 2020; Latonero, 2018). Esta aproximación se opone a la creencia que la legislación debe tomar un rol secundario a la autorregulación del mercado y que una regulación legal solo se traduzca en soluciones técnicas (Baldwin, Cave y Lodge, 2012). Así, por ejemplo, sistemas de IA que vulneran gravemente derechos humanos, como se ha postulado respecto de sistemas biométricos,¹⁸ deberían ser prohibidos y no secundar una autorregulación del mer-

18. Véase «Under Watchful Eyes: Biometrics, EU IT Systems and Fundamental Rights», *European Union Agency for Fundamental Rights*, 28 de marzo de 2018, disponible en <https://bit.ly/2RMiPMd>.

cado o una regulación técnica que simplemente clasifique a esta tecnología con una puntuación de riesgo.¹⁹

Independiente de qué tipo de regulación se escoja, no se debería hablar de una «definición legal de IA», ya que si bien existen muchas definiciones, ninguna cumple parámetros legales apropiados y cualquier intento definitorio podría ser limitante (Schuett, 2019), en consecuencia, lo óptimo es contar con un concepto de IA para ser considerado en ámbitos regulatorios que debe contemplar: 1) la descripción de sistemas sociotécnicos, ya que junto con el aspecto técnico se despliegan motivaciones propias de las personas que poseen una directa relación en su creación e implementación y dentro de un contexto social particular. La suma de dichos factores, inciden en los diversos impactos sociales adscritos (Hagerty y Rubinov, 2019) y 2) dicho concepto debe ser reflexivo, vale decir, entendiendo que un sistema puede ser usado para diversos propósitos y circunstancias.

¿Cómo discriminan los algoritmos?

Sobre la forma como discriminan los algoritmos, en particular los sistemas de toma de decisiones automatizadas, existen diversas investigaciones y estudios de carácter técnico. Es posible hacer un análisis mediante una clasificación establecida por Barocas y Selbst, complementada luego por Lerh y Ohm (Barocas y Selbst, 2016; Lehr y Ohm, 2017). Barocas y Selbst desarrollan un esquema sobre aquellas consideraciones que condicionan la generación de actos discriminatorios, siendo posible resumirlas en la generación de los llamados «sesgos» en los datos, algoritmos y modelos, los cuales pueden ser generados de forma accidental o intencional.

En materia legal «sesgo» se utiliza a menudo como sinónimo de prejuicio o un trato injusto, sin embargo, cuando se habla de sesgos en los modelos de ML se usa una definición estadística que describe un factor que cambia un resultado de cierta forma. Por su parte, otra definición de sesgos usada en ML se refiere a aquellos inconscientes emanados de las personas involucradas en el diseño de los sistemas o implícitos en los datos (Selbst, 2017).

Uno de los «sesgos en los datos» más conocidos es el de recopilación, el cual ocurre por la inferencia de sesgos incrustados en los mismos datos (Mehrabi y otros, 2019). En términos generales, los datos rara vez son neutros, están ligados a experiencias e historias de personas, por lo que reducirlos a modelos matemáticos sin tener en consideración las circunstancias que los rodean con el objeto de darle una aparente neutralidad, lleva ineludiblemente a resultados incompletos y equivocados.

19. Al respecto véase Amba Kak, Rashida Richardson y Roel Dobbe, «Submission To The European Commission On “White Paper On AI - A European Approach”», *AI Now*, 14 de junio de 2020, disponible en <https://bit.ly/3wLHYVQ>.

Otro de los sesgos más importantes relacionados con los datos, es el sesgo de representación, el cual puede manifestarse de varias formas. Generalmente, el sesgo de representación se presenta cuando se muestrea solo una parte de una población no siendo luego representativa de la totalidad a la cual el sistema es aplicado, subrepresentando usualmente a grupos vulnerables (Suresh y Guttag, 2019). Este sesgo es muy común en especial en los sistemas de toma de decisiones automatizadas, toda vez que estos son usualmente estandarizados, lo cual se explica por el reducido número de empresas que los desarrollan, no siendo viable técnica ni económicamente la creación de sistemas personalizados para cada usuario, por lo que en general son entrenados con el mismo conjunto de datos no representativos y usados a nivel mundial. Así, por ejemplo, los principales sistemas relacionados con decisiones en el área de la salud han sido entrenados con datos de tres estados norteamericanos (Kaushal, Altman y Langlotz, 2020) los cuales no siendo representativos en su propio país de origen, poseen un riesgo discriminatorio alto, el cual se incrementa y agrava en un contexto geográfico y social diferente (Creel y Hellman, 2021).²⁰

Respecto de sesgos en los algoritmos y modelos de ML, estos pueden manifestarse en decisión de diseñadores o como consecuencia de los propios límites técnicos de estos sistemas (Malik, 2020), toda vez que corresponden a modelos matemáticos de carácter cuantitativo, por lo que en definitiva hacen una reducción de datos, no siendo capaces de analizar todas las variables presentes en construcciones sociales (tales como peligrosidad, solvencia, enfermedad y también raza y sexo binario). Gran parte de las complejidades humanas que se pierden entre las características en los datos definidas y los resultados obtenidos por este tipo de sistemas.

Detrás de esta reducción, Catherine D'Ignazio y Lauren F. Klein hacen presente que hay una historia de desarrollo relacionado con técnicas heurísticas ampliamente establecidas en tecnología, cuyo fin es llegar a un resultado de forma rápida a través de atajos, lo que en definitiva se traduce en el sentido común. El problema es que el sentido común, tan usado en la política y en los discursos de poder, está cargado de sesgos, tal como lo han demostrado en sus múltiples estudios Daniel Kahneman y Amos Tversky (D'Ignazio y Klein, 2020, cap. 1; Tversky y Kahneman, 1974).

De la misma forma, estas autoras señalan que los principales tipos de algoritmos de ML, como los de regresión lineal, fueron desarrollados por los eugenistas del siglo XX²¹ y uno de los algoritmos más usados en ML, de agrupamiento jerárquico o clúster, deriva de un modelo de segregación habitacional racista de los años cincuenta

20. Para más antecedentes sobre la clasificación de sesgos véase Barocas y Selbst (2016), Zemel y otros (2013) y Lehr y Ohm (2017).

21. El eugenismo es un movimiento político e ideológico que predica una mejora cualitativa, biológica, «natural» de la población, frente a la mejora moral, económica o «cultural». Véase Julio Pérez, «Eugenismo», *Apuntes De Demografía*, 15 de junio de 2010, disponible en <https://bit.ly/3wOkfER>.

(D'Ignazio y Klein, 2020).²² En particular, estos algoritmos tienen una tendencia o preferencia hacia la optimización de un resultado predefinido y, en consecuencia, generan como contrapartida exclusión.

Asimismo, es importante hacer presente que cuando se habla de «predicción» nos referimos a un resultado de correlaciones y no a una predestinación (Malik, 2020). Las correlaciones como predicciones son sensibles a los cambios de contextos y, por lo tanto, imprecisas. Definir el resultado de correlaciones como predicciones generalizables es extrapolar un resultado fijo que solo es dado para un momento y tiempo determinado, lo cual evidentemente sería ambiguo y limitado (Lipton, 2018).

Estas correlaciones llamadas espurias, sin ninguna causa directa, se basan en constructos o abstracciones, denominados *proxies*, que intentan traducir conceptos complejos en elementos medibles. La delincuencia, la solvencia, la salud, son abstracciones no elementos físicos, por lo que el análisis medible de estas construcciones sociales por parte de los ML se conciben a través de dichos *proxies*, los cuales se traducen por ejemplo en, tasas de *clicks*, códigos postales, cupos de tarjetas de crédito o amigos de Facebook (Malik, 2020). Con frecuencia estos *proxies*, de forma accidental o intencional, se corresponden con marcadores de características protegidas (determinados *clicks* pueden marcar una orientación sexual erróneamente, así como un código postal puede marcar una nacionalidad, condición social o etnia equivocada), en consecuencia, además de imprecisos, el uso de *proxies* fácilmente conlleva arbitrariedades y discriminación.

Sin perjuicio de lo anterior, explicar cómo se produce la discriminación a través de sistemas de IA exclusivamente desde el punto de vista técnico sería una limitación, puesto que como se ha señalado, la IA es un concepto «sociotécnico» que solo se explica teniendo en consideración propósitos, motivaciones y relaciones sociales que influyen en su desarrollo e implementación. En consecuencia, siguiendo la visión de teorías decoloniales (Lander y Castro-Gómez, 2000) y feministas (Mendez, 2015), este análisis debe ser complementado con un cuestionamiento crítico de los impactos negativos de la IA como una manifestación de poder (VeneKlasen y otros, 2002) presente en todas las sociedades (Costanza-Chock, 2020; D'Ignazio y Klein, 2020; Benjamin, 2020).

En efecto, a lo largo de los siglos todos los grandes conflictos sociales han estado inexorablemente supeditados a una dinámica de poder y distribución de privilegios de unos sobre otros, que históricamente han generado grupos oprimidos, en constante desventaja, sujetos a discriminaciones y exclusiones. Lo anterior se traduce en que nuestros sistemas sociales y políticos son profundamente injustos y desiguales (Quijano, 2000; Collins, 2002; Freire, 1970).

22. Para más información al respecto véase Catherine Muñoz, «¿La tendencia a generar vínculos con personas similares es realmente parte de la naturaleza humana?», *El Periodista*, 12 de noviembre de 2020, disponible en <https://bit.ly/3fSDiHb>.

En concordancia con el punto anterior, una explicación simplemente técnica se transforma en incompleta, ya que lleva a un análisis casuístico y de actos independientes, lo cual ha sido criticado por diversos autores (D'Ignazio y Klein, 2020; Benjamin, 2020), puesto que, para una acabada comprensión, se debe entender el problema de la discriminación y los sesgos discriminatorios como una situación estructural, pues así es como ocurre en la sociedad (Young, 2002).

Por su parte, la visión que generalmente se tiene de estas tecnologías es de neutralidad u objetividad. Sin embargo, desde la perspectiva del conocimiento situado (Haraway, 2013), se comprende que la IA en ningún caso es neutra, poseyendo en la práctica una aparente objetividad, siempre desde un punto de vista determinado. En ese sentido, tal como Donna Haraway lo describió respecto de la ciencia, los desarrollos tecnológicos están situados en un contexto donde ha existido históricamente una brecha entre hombres y mujeres, además de exclusión de grupos minoritarios, por lo que esta tecnología se diseña y desarrolla desde una perspectiva particular, la de hombres blancos de elite (Benjamin, 2020), ya que son ellos los que ocupan las posiciones más altas y privilegiadas dentro del campo de los desarrollos tecnológicos (D'Ignazio y Klein, 2020, cap. 1).

Adicionalmente, los sistemas de toma de decisiones automatizadas, cuando están destinados a resolver problemas sociales, son generalmente de carácter punitivo, verbigracia, predictores de riesgos, de delitos o de fraudes, llamando la atención que no se implementen sistemas que tengan fines constructivos, como la rehabilitación o la reducción de desigualdades (Birhane, 2020). A lo anterior se debe agregar que dichos sistemas son dirigidos casi exclusivamente a grupos marginados y vulnerables, pudiendo concluir que su afectación puede estar prácticamente asegurada, así como que su implementación puede transformarse en una forma de dominación, poder o vigilancia encubierta, tal cual funcionan otras herramientas dotadas de IA como la vigilancia masiva (Yu, 2020). Cobraría lógica, *a contrario sensu*, la inexistencia de sistemas predictivos de fraudes a nivel empresarial o de elite o sistemas predictivos de violencia, delitos o abuso en poblaciones de estratos sociales altos.

Por su parte, cuando los sistemas de IA no benefician a las estructuras de poder, estas los ponen en duda. Un ejemplo se recoge en un reciente artículo (Malik, 2020) donde se cita un caso (Iwanicz-Drozdowska y Witkowski, 2016) que describe la respuesta del sector crediticio estadounidense a un Informe de la Oficina de Protección Financiera del Consumidor, el cual detectó discriminaciones en el otorgamiento de créditos mediante una investigación que usó una metodología de IA sobre inferencia de rasgos no observados mediante *proxies*. El sector crediticio argumentó que la metodología usada para acusarlos era «una mera conjetura con apariencia sofisticada», siendo que estas son las mismas metodologías usadas para decidir este tipo de créditos.

Finalmente, un ejemplo de un uso derechamente intencional de este tipo de siste-

mas para propósitos discriminatorios fue conocido a través de un estudio realizado en los Estados Unidos²³ que demostró la existencia de actos discriminatorios como una parte integrante del modelo de negocios de publicidad selectiva de Facebook. Los perfiles creados por inferencia de datos o afinidades que se construyen para cada usuario permiten agruparlos en diversas categorías, que luego Facebook utiliza para ofrecer a sus diversos anunciantes una herramienta que les da la posibilidad de excluir a grupos enteros de acuerdo con las características específicas de los perfiles que reúne (afroamericanos, inmigrantes, orientación sexual, identidad de género, clase social, entre otros), dentro de la categoría «demografía». Esta herramienta impedía que estos grupos de personas tuviesen acceso a diversos anuncios de viviendas en una abierta discriminación, expresamente prohibida de forma legal en los Estados Unidos, debido a una histórica discriminación residencial de carácter racista existente hasta el día de hoy en dicho país.²⁴

Los desafíos de las normas de no discriminación frente a una discriminación algorítmica

Análisis comparativo situado en Estados Unidos y Europa

En términos generales las normas sobre no discriminación, sin distinción de su procedencia, protegen a las personas, específicamente a aquellas cuyas características las exponen a una probabilidad significativa de estar en desventaja y sufrir menoscabo (Khaitan, 2015); este es el criterio para determinar las denominadas «categorías sospechosas», que describen una pertenencia a grupos en desventaja que merecen una protección reforzada. No obstante, no existe un único concepto de discriminación, ni un único mecanismo establecido para determinar cuándo un acto es discriminatorio, dependiendo en gran medida de la cultura y contexto social.

En efecto, el alcance de las normas de no discriminación está asociado a su contexto histórico y social. De este modo, las normas específicas europeas, que son de reciente data,²⁵ tomaron como base leyes norteamericanas sobre no discriminación

23. Julia Angwin, Ariana Tobin y Madeleine Varner, «Facebook (Still) Letting Housing Advertisers Exclude Users By Race», *ProPublica*, 21 de noviembre de 2017, disponible en <https://bit.ly/3i3W8hk>.

24. Julia Angwin, Noam Scheiber y Ariana Tobin, «Dozens of Companies Are Using Facebook to Exclude Older Workers from Job Ads», *ProPublica*, 20 de diciembre de 2017, disponible en <https://bit.ly/2Ti5xqX>. Esta investigación destaca la amplitud de este problema, Facebook tiene el 22% del mercado de anuncios digitales en los Estados Unidos y se ha demostrado que la publicidad selectiva es su principal modelo de negocios (*digital redlining*), el cual se centra en prácticas tecnológicas que consolidan actos discriminatorios contra grupos ya segregados por la sociedad.

25. En el derecho europeo el principio de no discriminación se encuentra consagrado en dos de las principales normas comunitarias, el Consejo de Europa y el Convenio Europeo de Derechos Humanos, y además existen normas específicas, tales como, la Directiva 2000/43 del Consejo de Europa, relativa a

a discapacitados, siendo situadas y contextualizadas desde su perspectiva cultural. En ese sentido y desde el punto de vista cultural, las normas estadounidenses tendrían una connotación sobre la promoción de libertad y elección individual desde una perspectiva de derechos civiles, mientras que las normas europeas se basarían en un carácter de bienestar colectivo desde la justicia social (De Búrca, 2012).

La primera aproximación en temas de discriminaciones es determinar qué se entiende por igualdad. La igualdad es un concepto jurídico que se establece sobre la noción aristotélica de tratar igual a los iguales y desigual a los desiguales (Díaz de Valdés y Manuel, 2015). Aquella fórmula básica, presente en múltiples tratados internacionales, constituciones y normas, es conocida como igualdad ante la ley o igualdad formal, una primera aproximación en la que coinciden gran parte de las legislaciones (Khaitan, 2015).

Dicha definición de igualdad formal necesita ser complementada con estándares externos de justicia para que revele un carácter sustantivo. La primera interpretación de estándar de justicia indica que la igualdad es una garantía de no discriminación hacia grupos vulnerables; es esta la forma como generalmente se conoce y guarda relación con la protección a personas pertenecientes a los referidos grupos protegidos, cuyo análisis se hace a través de una «evaluación de razonabilidad», en el caso de Europa, y un «escrutinio escalonado», en el caso de Estados Unidos (Ronconi, 2018).

Siguiendo con el punto anterior, el concepto clásico de discriminación se define de manera comparativa en los siguientes términos:

Un trato diferente de individuos o grupos basado en criterios arbitrarios adscriptivos o adquiridos, como sexo, raza, religión, edad, estado civil o parental, discapacidad, orientación sexual, opiniones políticas, antecedentes socioeconómicos, afiliación y actividades sindicales.²⁶

La discriminación puede ser directa, siendo aquella fácilmente detectable y ocurre cuando «una persona es tratada menos favorablemente que otra por motivos de sexo y género, edad, nacionalidad, raza, etnia, religión o creencia, salud, discapacidad, orientación sexual o identidad de género, de lo que es, ha sido o sería tratada

la aplicación del principio de igualdad de trato de las personas independientemente de su origen racial o étnico, y la Directiva 2000/78 del Consejo de Europa, relativa al establecimiento de un nuevo marco general para la igualdad de trato en el empleo y la ocupación. Estas normas, como todas las normas comunitarias, tienen prioridad sobre el derecho interno y han sido la base para la promulgación de normas nacionales en la mayor parte de países de la Unión.

26. Véase «Trabajadores indocumentados. Estados Unidos de América», CIDH, Informe 50/16, caso 12.834, párr. 75, del 30 de noviembre de 2016; y «El trabajo, la educación y los recursos de las mujeres: La ruta hacia la igualdad en la garantía de los derechos económicos, sociales y culturales», CIDH, OEA/Ser.L/V/II.143 Doc. 59, párr. 16, del 3 de noviembre de 2011, disponible en <https://bit.ly/3gIe2EZ>.

otra persona en una situación comparable».²⁷ En Estados Unidos este tipo de discriminación se denomina tratamiento dispar (*disparate treatment*).

De la misma forma, la discriminación también puede ser indirecta, la cual ocurre cuando una disposición o práctica aparentemente neutra pone a las personas que comparten una característica protegida en una desventaja desproporcionada en comparación con otras personas (Kitching, 2005). Anuncios de trabajo, de viviendas, el caso Ofqual sobre asignación de calificaciones, son un tipo de discriminación indirecta. Este tipo de discriminaciones en la legislación norteamericana se denomina impacto dispar (*disparate impact*) y es la regla general en materia de discriminación algorítmica.

Hasta esta clasificación no hay diferencias sustanciales entre la aproximación europea y norteamericana. Las diferencias se aprecian al momento de evaluar o determinar cuándo un acto es discriminatorio. A continuación, se exponen dos sistemas implementados en cada jurisdicción para graficar estas diferencias:

Caso 1, Reino Unido

Joint Council for the Welfare of Immigrants (JCWI), una organización no gubernamental de apoyo a inmigrantes en el Reino Unido, dedujo una demanda en el mes de junio de 2020 contra el Estado por discriminación directa por motivos de raza sobre un grupo de personas migrantes, en violación de la ley de igualdad,²⁸ norma interna en este país.²⁹

Según relata en su demanda, desde el año 2015 el Ministerio del Interior del Reino Unido ha utilizado un modelo de riesgo predictivo de ML, Streaming Tool, para evaluar las solicitudes de visa que recibe y clasificarlas en diferentes categorías de riesgo: roja, ámbar o verde. Se llegó a la conclusión de que la nacionalidad de los solicitantes era un factor significativo para la evaluación de riesgo. Los solicitantes de nacionalidades «sospechosas» en igualdad de condiciones que personas de otras nacionalidades, tenían una alta incidencia en recibir una calificación de riesgo alta, a pesar de que consideraciones por nacionalidad para el otorgamiento de visas están prohibidas en dicho país.

En este caso se retrata una discriminación directa contra grupos de población migrante. La discriminación directa se considera una excepción en materia de discriminación algorítmica, sin embargo, según el criterio de las víctimas, este sistema

27. Véase «Direct Discrimination», *European Institute For Gender Equality*, 2017, disponible en <https://bit.ly/3wK4wqg>.

28. Véase Equality Act 2010, disponible en <https://bit.ly/3vJrNI3>.

29. Para más detalles sobre los documentos legales presentados en este caso véase «How We Got the Government to Scrap the Visa Streaming Algorithm - Some Key Legal Documents», *Foxglove*, 4 de agosto de 2020, disponible en <https://bit.ly/3c3vxNq>.

fue implementado con la intención directa de discriminar, usando un mecanismo de refuerzo negativo mediante eventos irracionales e injustos y mecanismos que desincentivaban que una persona tomara una decisión diferente al sistema, puesto que conllevaba una importante carga laboral adicional.

Los tribunales europeos realizan un examen de «razonabilidad» compuesto de varios pasos para evaluar un acto discriminatorio. En primer lugar, se debe determinar si se ha perseguido una finalidad legítima constitucionalmente no prohibida, luego, si no existió algún medio alternativo para lograr el mismo fin y, finalmente, se realiza un análisis sobre proporcionalidad (Ronconi, 2018; Khaitan, 2015).

En este caso en particular, se probó la discriminación utilizando un comparador, esto es, personas con las mismas condiciones pero de otra nacionalidad no asociada a migración, y se hizo un análisis de razonabilidad destacando aquellos puntos por los que se consideró que el funcionamiento del sistema era irracional, particularmente, se hizo presente la existencia de un sistema de refuerzo que se «alimentaba de eventos» irracionalmente injustos, a saber, un evento negativo podría ser una visa vencida, el número de personas que solicitan visa dentro de un tiempo establecido superando cierto umbral o un rechazo anterior, entre otras situaciones que no guardan ninguna racionalidad para ser antecedente posterior de denegación de visas. Luego, a través de pruebas estadísticas se pudo comprobar que el mecanismo de refuerzo del sistema basado en los efectos negativos anteriores imponía condiciones arbitrarias en el análisis de futuras visas para personas de determinados países.³⁰

Caso 2, Estados Unidos

Allegheny Family Screening Tool (AFST) es un modelo estadístico predictivo de abuso infantil que elabora puntajes sociales de riesgo, de 1 a 20, siendo este último el riesgo más alto. Si el puntaje es alto se activa un sistema de investigación humana. El sistema no predice directamente el abuso infantil, sino que como ya se ha señalado, usa *proxies* o marcadores que se consideran asociados al mismo, así, por ejemplo, uno de los marcadores iniciales fue el número de llamadas de denuncias, lo cual es resultado un *proxy* complejo puesto que existe un sesgo en denuncias respecto de familias afrodescendientes tres veces mayor al de familias blancas.

Asimismo, el sistema produce resultados básicamente en función de los datos pú-

30. Más detalles sobre este caso véase «We Won! Home Office to Stop Using Racist Visa Algorithm», *Joint Council for The Welfare of Immigrants*, 2020, disponible en <https://bit.ly/2TtJw8T>. Véase también Henry McDonald, «AI System for Granting UK Visas is Biased, Rights Groups Claim», *The Guardian*, 29 de octubre de 2019, disponible en <https://bit.ly/3yPjj4z>. Luego de la presentación de la demanda, y antes que se iniciara el proceso judicial, finalmente el Gobierno del Reino Unido decidió, con fecha 3 de octubre de 2020, retirar el sistema del proceso automatizados de visas, el cual pasó a ser llevado exclusivamente por personas.

blicos disponibles, los cuales son en su gran mayoría de personas pobres extraídos de sistemas de asistencia médica pública, sistemas de beneficios sociales, seguros de ingresos suplementarios, entre otros, no teniendo como contrapartida, acceso a datos sobre personas que acceden a servicios privados, es decir, no se cuenta con los datos de aquellas personas que por ejemplo siguen tratamientos de drogas o médicos en establecimientos privados y pueden ser indicadores de abuso infantil. Adicionalmente, según estudios, el porcentaje de predicción correcta ha sido bajo, existiendo muchos falsos positivos y negativos (Eubanks, 2018, cap. 4).

En definitiva, gran parte de las variables de este sistema de riesgo son medidas directas de pobreza, en otras palabras, el sistema tiende a asociar abusos infantiles a pobreza, lo cual es un acto discriminatorio, ya que no analiza un comportamiento concreto, sino que se basa en una condición social, además de atentar contra la dignidad y privacidad de las personas, la salud e incluso la vida (Gilman, 2020). En la ciudad de Illinois se prohibió el uso de un sistema similar³¹ luego de que miles de niños fueran erróneamente asignados con alta probabilidad de muerte o lesiones, mientras que muchos niños que murieron o terminaron siendo abusados no fueron alertados a tiempo.

En este segundo ejemplo, se presenta un sistema para el bien social en Estados Unidos, la prevención del abuso infantil, que es aparentemente neutro, al igual que el sistema que determinó los puntajes de acceso universitario en el Reino Unido, pero que aplicado a la generalidad de la población pone a determinados grupos de personas que comparten una característica protegida en una situación de desventaja, pobreza y racismo, evidenciándose en consecuencia, una discriminación indirecta o de impacto dispar (Suresh y Guttag, 2020; Friedler y otros, 2019).

Si bien estos sistemas y sus efectos son reales, en el caso del primero aún se sigue implementando. Para efectos ilustrativos, no tomaremos un caso judicial en particular, sino que ejemplificaremos hipotéticamente cómo se realizaría una evaluación para determinar discriminación en este tipo de sistemas en los tribunales norteamericanos. La Corte Suprema de Estados Unidos distingue «niveles de escrutinio» de mayor o menor exigencia dependiendo del rasgo protegido afectado (Kitcing, 2005). Al contrario de lo que ocurre en Europa y en la mayor parte de Latinoamérica, en Estados Unidos existe un listado limitado de rasgos protegidos, algunos de rango constitucional, como la raza y el sexo, y otros de categoría legal, como la discapacidad, siendo aplicadas distintas normas según el acto lo cometa el Estado (norma constitucional) o un particular (leyes especiales) (Hellman, 2020a).

En este ejemplo, como es un sistema público, correspondería aplicar la norma constitucional. Si determinamos que la discriminación finalmente guarda relación

31. Gary Marx y David Jackson, «Data Mining Program Designed To Predict Child Abuse Proves Unreliable, DCFS Says», *Chicago Tribune*, 6 de diciembre de 2017, disponible en <https://bit.ly/3vzYbgV>.

con un tema de «raza» que se asocia con la clase social, se aplicaría un escrutinio estricto, por el cual se determina un acto discriminatorio sin considerar otras circunstancias como intenciones o proporciones, salvo un interés gubernamental importante, solo siendo necesario probar una razón racista explícita. Si no hay una razón racial explícita, si solo se alega que se discrimina a las personas por una condición social, se aplica un nivel de escrutinio menos estricto y se toman en consideración factores como la intención y otras circunstancias a la hora de evaluar una discriminación algorítmica (Hellman, 2020b). Por el contrario, en Europa frente a una discriminación algorítmica indirecta se actúa de la misma forma independientemente del grupo o característica afectada, es decir, se hace un mismo análisis racional y no se considera en estos casos la existencia de una intención en la práctica discriminatoria.

Dificultades y desafíos que enfrentan las actuales normas de No Discriminación

Como se ha señalado, la discriminación como regla general se prueba a través de un *comparador*. En el primer caso mencionado (Streaming Tool) pudiese ser menos complejo establecer un comparador (una persona de otra nacionalidad en las mismas condiciones), pero en la generalidad de las discriminaciones indirectas encontrar un comparador suele ser una tarea casi imposible, ya que en estos casos las comparaciones son a nivel de grupos de personas en circunstancias muy específicas (Farkas, 2011).

Adicionalmente, algunos autores han señalado que el concepto de *características protegidas* puede ponerse en tela de juicio y también pueden surgir nuevos y contraintuitivos *proxies* que pueden ser marcadores de clases protegidas difíciles de detectar (Wachter, Mittelstadt y Russell, 2020) Mittelstadt, y Russell 2020.

Por otra parte, es importante destacar que, dentro de una discriminación mediante sistemas de IA, tanto directa como indirecta, una persona puede ser afectada por error, no perteneciendo a un grupo protegido, pero siendo asociada a alguno. Así, por ejemplo, en el caso de los perfiles de usuarios creados por Facebook, muchos de ellos son «perfiles por afinidad», es decir, no se infieren directamente sobre datos personales sino que agrupan a determinadas personas de acuerdo con intereses asumidos, por lo que poseen altas tasas de imprecisión situación que no estaría contemplada en las actuales normas de no discriminación ni europeas ni norteamericanas (Wachter, 2020)

Otro de los grandes problemas es la dificultad que representa para una víctima probar un caso, en primer término, lograr obtener una discriminación *prima facie*, que es una descripción de un caso al inicio de un juicio acompañado de un nivel de evidencia tal que permita que un tribunal traslade la carga probatoria al acusado (Khaitan, 2015). Esto es esencial sobre todo en una discriminación algorítmica, dada la desventaja mayor en la que se encuentran las víctimas a la hora de aportar pruebas específicas de estos sistemas de características complejas y opacas.

Por su parte, la prueba estadística usada en los casos de discriminación por sistemas de IA, tanto en Europa como Norteamérica, es la de mayor relevancia, ya que se trata de un medio útil cuando disponer de pruebas es un desafío, y además permite exponer desigualdades estructurales ocultas respecto de otras personas (Goodman y Flaxman, 2017). Sin embargo, estas pruebas tienen importantes limitaciones y complejidades, ya que, si bien se han creado métricas y sistemas para esta finalidad, estas dependen de lo que se crea correcto por punto de equidad o justicia, y es aquí donde cobra relevancia el concepto de *fairness* y lo que representa.

Las métricas de *fairness* norteamericanas funcionan bajo cierta lógica netamente matemática (Mehrabi y otros, 2019; Xiang y Raji, 2019b; Green y Hu, 2018; Hu y Kohler-Hausmann, 2020) que no es completamente aplicable a los niveles de razonamiento subjetivo europeo o latinoamericano, donde aún no existe consenso sobre su equivalente jurídico en español. Lo anterior es de suma importancia, puesto que distintos tipos de sistemas o métricas pueden, para un mismo caso, dar resultados completamente diferentes (Hellman, 2020a). Por la importancia y extensión de esta temática particular, corresponde desarrollarla en un estudio independiente.

En conexión con el punto anterior, uno de los mayores problemas se traduce en una falta de transparencia, pero no aquella entendida como el conocimiento completo en detalle de un sistema, sino en términos de contar con información relevante para probar un caso, como antecedentes sobre decisiones tomadas por las personas detrás del desarrollo e implementación de un sistema, las razones de establecimiento de determinados *proxies si fuese el caso*, en general, información estratégica sobre como el sistema produce resultados discriminatorios, que pueda ser complementada con pruebas estadísticas. No obstante, existen casos en los que puede ser necesario además saber el detalle de componentes de estos sistemas, por ejemplo, en el caso que se requiera analizar un determinado *dataset* para corroborar sesgos discriminatorios en ellos. En tales circunstancias, pueden existir barreras que se relacionan con derechos de tipo industriales, como los secretos comerciales.

Teoría conceptual sobre discriminación algorítmica situada en América Latina

Como ya se ha señalado, la regulación sobre discriminación está situada dentro de un contexto determinado y, en consecuencia, las normas norteamericanas y europeas tienen distintas aproximaciones y distintas formas de determinar una discriminación algorítmica, siendo una materia esencialmente social. Siguiendo el punto anterior, desde el contexto latinoamericano, si bien contamos con una historia común con el derecho europeo, existen matices legales propios de cada país.

De la misma manera, desde el punto de vista técnico, gran parte de estas investigaciones, estándares y procedimientos sobre sistemas y métricas de detección y

mitigación de sesgos, que se usan tanto para prevenir sesgos como para probar un caso en un juicio, se han realizado en Estados Unidos centrados en la detección de discriminación algorítmica relacionada, principalmente, con el racismo contra afrodescendientes (Mehrabi y otros, 2019; Selbst, 2017; Hanna y otros, 2020) pudiendo estas herramientas no ser del todo aplicables para otras regiones como Latinoamérica, donde los marcadores racistas pueden ser diferentes.

Así, por ejemplo, un mecanismo para detectar discriminación considerando el tono de piel usado como marcador de racismo, hipotéticamente puede no ser capaz de detectar discriminación en una población donde el tono de piel está compuesto por matices (Mehrabi y otros, 2019). Lo anterior significa no solo que no sería útil, sino que podría generar un resultado de «falso negativo» de discriminación, por lo que se considera primordial el desarrollo de estudios técnicos y legislativos regionales sobre este tema.

En vista de lo anterior, se desarrolla a continuación una caracterización de aquellos elementos distintivos y propios de Latinoamérica en relación con actos discriminatorios que pueden ser relevantes a la hora de evaluar una discriminación algorítmica, pudiendo ser tomados en consideración para futuras evaluaciones legales y técnicas. Para efectos prácticos este análisis se centra en aspectos discriminadores relacionados con el racismo.

Racismo y grupos protegidos latinoamericanos desde la perspectiva interseccional

Si bien los grupos protegidos en América Latina están determinados por tratados internacionales³² y legislación interna, cobra relevancia en la región su análisis desde una perspectiva interseccional.³³ Para lo anterior, debemos considerar el origen histórico latinoamericano para advertir que la raza como construcción social y la clase (asimismo el género) están profundamente entrelazados, pudiendo en una evaluación algorítmica ser difíciles de detectar o determinar la característica o grupo que se está afectando de forma principal, siendo generalmente la acumulación de varios de ellos. Esta unión de categorías sospechosas tiene su origen en dos eventos históricos ocurridos simultáneamente, la constitución de América y la creación del capitalismo colonial. La clasificación social de la población colonizada sobre la idea de raza sirvió, en primera instancia, de justificación para la nueva dominación eu-

32. Principalmente por la Convención Americana sobre Derechos Humanos (o Pacto de San José, adoptado en 1969).

33. La interseccionalidad corresponde a una afirmación que considera que la etnia, la clase, el género, la sexualidad, la capacidad, entre otras categorías, no operan de forma independiente y excluyente, sino que se superponen y potencian construyendo complejos sociales de múltiples capas que explican y dan forma a desigualdades estructurales presentes en la sociedad (Hill Collins, 2002).

ropea, practica denominada colonialidad (Mignolo, 2003). Esta distinción produjo nuevas identidades sociales (indios, negros, mestizos) y redefinió conceptos como español o portugués, de simples indicadores geográficos a indicadores jerárquicos de superioridad humana. Esta clasificación por rasgos fenotípicos sirvió, a su vez, para el control y explotación sobre el trabajo, dando paso al desarrollo del reciente capitalismo. Entonces, la clase social y raza quedaron unidas de forma estructural desde su origen (Quijano, 2000) hasta hoy. Estos tipos de discriminaciones, junto con la discriminación por motivo de género tienen un origen estructural, ya que son productos de circunstancias sociales históricas concatenadas, no siendo en ningún caso eventos independientes y esporádicos.

Otra característica de la región, especialmente en los países del pacífico sur, es que las naciones se fueron construyendo socialmente en base a una población mayoritariamente mestiza, quedando compuesta en menor medida por europeos e indígenas, sin embargo, las estructuras jerárquicas racistas que ahora sostenían una nueva economía fueron mantenidas socialmente (Silva y otros, 2018).

El racismo latinoamericano: el daltonismo racial

En países como México (Moreno Figueroa y Saldívar Tanaka, 2016), Colombia, Argentina, Chile y Uruguay, existe una arraigada creencia de superación de los problemas raciales basados en la existencia de un generalizado mestizaje, sin embargo, como se ha mencionado en el párrafo anterior, en estas sociedades se han mantenido las estructuras jerárquicas fundadas en el racismo (Wade, 2017).

Esta disonancia, denominada «daltonismo racial», se caracteriza porque las personas creen que desigualdades, exclusiones y discriminaciones no son un problema de racismo, viéndolo como un problema pasado y explicando cualquier tipo de injusticia desde otras perspectivas, tales como consecuencias del mercado, la naturaleza o la suerte (Bonilla-Silva y Embrick, 2006). El daltonismo racial puede dificultar una detección de discriminación algorítmica, ya que es probable que una persona no pueda percibir por sí misma que está siendo discriminada por motivos raciales, pudiendo atribuir determinadas situaciones discriminatorias a hechos injustos de la naturaleza o el azar, ya que existe una tendencia en ciertos países latinoamericanos a autoclasificarse como blancos (Lizcano Fernández, 2005).

En contrapartida, esta ceguera de tonalidades exagera aquellas diferencias que son evidentes, representada en una abierta y aceptada exclusión y discriminación de comunidades indígenas, que actualmente se extiende hacia la nueva población migrante. Las comunidades de pueblos originarios en América Latina han sido fuertemente estereotipadas, asociándolas a determinadas clasificaciones morales (Espinosa y Cueto, 2014).

La académica Safiya Noble ha estudiado la construcción de estos estereotipos en

los motores de búsquedas de internet, en cuyos resultados es posible apreciar un sistema jerárquico de privilegio sobre la base de fenotipos, donde las identidades raciales son construidas por «etiquetas» impuestas por un grupo superior (Noble, 2018: 82).

Siguiendo a esta académica, si buscamos en Google imágenes, «hombres chilenos», el resultado que arroja muestra a hombres blancos, sonriendo y posando (no reflejando al común de los hombres chilenos) sin ningún contexto en particular. No aparecen alusiones a la cultura, tradición e historia y los botones de búsqueda adicional sugerida son triviales (jóvenes, famosos, normales, actores).

Si luego buscamos, «hombres mapuches», el resultado de imágenes es diferente, en primer lugar, no aparecen solo hombres (aparecen niños y mujeres también), por lo que no se trata de una búsqueda precisa. La mayor parte de las imágenes representan a personas enojadas o confrontaciones caracterizadas por sus vestimentas y dentro de un contexto cultural exacerbado, los botones de búsqueda adicional sugerida están contextualizado como vestimentas, historia, accesorios mapuches, tradición, guerra.

Al comparar ambos resultados, los hombres mapuches (además de mujeres y niños) son sobrerrepresentados, generándose un estereotipo, personas enojadas, «serias», que se desenvuelven en un contexto cultural determinado alejado de la modernidad. La sobrerrepresentación, como el modo de catalogar y definir a un grupo por parte de otro dominante, por sí mismo constituye un acto discriminatorio (Wieviorka, 2002).

En resumen, una discriminación derivada del uso de un sistema de IA en Latinoamérica posee particularidades propias de la región: 1) en primer lugar, las características sospechosas de raza y clase, así como género, están unidas por complejas capas interseccionales y estructurales, por lo que, además de ocultar al racismo como un elemento cultural relevante, frente a una evaluación de sesgo algorítmico puede ser difícil y confuso determinar qué categoría está siendo afectada; 2) la existencia del «daltonismo racial» dificulta que las propias personas sean conscientes de que están siendo discriminadas, cuestión, por lo demás, arduamente comprobable debido a que muchas prácticas discriminatorias son socialmente aceptadas; 3) los marcadores *proxies* y parámetros para determinar sesgos discriminatorios en sistemas de IA que se han desarrollado en países del norte global, no son compatibles con la realidad latinoamericana, pudiendo distinguirse marcadores regionales diferentes y propios, entre otros, tonalidades de piel, apellidos, consideraciones fenotípicas como altura, rasgos de ojos, pelo, forma de nariz y, al encontrarse interconectados con la clase, los marcadores como código postal, colegio, nivel de estudios, actividad agrícola son, a su vez, marcadores racistas (Salgado y Castillo, 2018); 4) existe una cierta aceptación social de prácticas discriminatorias respecto de pueblos originarios, lo que también es un elemento característico asociado a las contemporáneas migraciones.



Figura 1. «Hombres chilenos». Búsqueda con Google el 12 de marzo de 2021.



Figura 2. «Hombres mapuches». Búsqueda con Google el 12 de marzo de 2021.

Rescatando la racionalidad y colectividad latinoamericana

Principio relacional: volvamos a ser puentes

Los pueblos originarios en América Latina poseen en común una cosmovisión relacional, que comprende una vivencia en comunidad contraria a todo concepto de individualidad y competencia impuestas desde tiempos coloniales, definiéndose esencialmente como sujetos relacionales (Meza Salcedo, 2017). A pesar de diferencias de lenguajes y culturas, este principio es transversalmente característico y compar-

tido también con pueblos de la lejana África (Mhlambi, 2020). Trabajar y vivir en comunidad, de forma solidaria, con ayuda mutua y horizontalización de las familias, caracteriza a la totalidad de los pueblos originarios de la región (Estermann, 2008). Este artículo propone rescatar este principio filosófico propio de la región que puede servir de fuente para reinterpretar los principios de igualdad y el concepto de discriminación y ser una base para una nueva construcción relacional como sociedad y de cara a la tecnología.

Al respecto, la Comisión Interamericana de Derechos Humanos³⁴ recientemente ha declarado que los derechos humanos no son conceptos estáticos, debiendo ser reinterpretados constantemente a la luz de los cambios sociales, tomando como regla de interpretación principal «la opción más favorable a la persona», así como la evolución de los derechos fundamentales en el derecho internacional contemporáneo.

La racionalidad nos invita a dejar de ser meros observadores de nuestro entorno y pasar a ser participantes de este, reencontrarnos con el complejo conocimiento ancestral, sintiéndonos a salvo en comunidad con personas y medio ambiente. Desde esa perspectiva, podemos llegar a la conclusión de que no todo es solucionable tecnológicamente (Escobar, 2019), lo cual no significa un rechazo, sino un empoderamiento de decisiones donde se privilegie la cocreación de diseños tecnológicos en beneficio del buen vivir comunitario y de acuerdo al respeto por los más débiles.

Según Francisco Varela la relacionalidad nos invita a dejar de evadir las problemáticas sociales y pasar a ser parte de ellas, construyendo nuestra propia realidad (Varela, Thompson y Rosch, 2017). Desde esa perspectiva el uso generalizado de predicciones no tendría sentido, ya que preponderaría la reflexión y el aprendizaje en conjunto (Birhane, 2021). Desde el principio relacional es posible generar un poder contrarrestante para mitigar los actos discriminatorios, tomando decisiones sobre qué tecnología se requiere y cuáles no son convenientes para oponerse a aquellas que se consideran riesgosas. Es posible elegir el desarrollo de innovación local e inclusiva incorporando el contexto y experiencia de los grupos más vulnerables, generando, en definitiva, una sociedad más justa con nuestras decisiones.

Igualdad como no dominación desde una perspectiva colectiva

Asimismo, es necesario hacer presente que la igualdad debe también ser interpretada desde una visión colectiva. Al respecto, siguiendo las reflexiones de Iris Marion Young, se establece que es una aproximación errada aquella que naturalmente ha considerado la igualdad como principio individual, es decir, cuya evaluación se hace a partir de los individuos (Young, 2001).

34. «Compendio Igualdad y Discriminación: Estándares Interamericanos», OEA/Ser.L/V/II.171Doc. 31, CIDH, del 12 de febrero de 2019, disponible en <https://bit.ly/3vBfoIn>.

En efecto, desde la perspectiva individual no es posible detectar ni evaluar injusticias, ya que se generan juicios sobre obligaciones o suficiencia (una persona no tiene para comer mientras a otras les sobra), pero no se toman en consideración circunstancias o aspectos que unen a estas personas con un grupo específico, siendo posible determinar que un hecho es desigual (una persona pobre versus una persona rica) mas no injusto. En cambio, para determinar que algo es injusto o no, discriminatorio o no, necesariamente se debe hacer una evaluación a nivel de grupos. Desde esta visión colectiva es posible determinar que las desigualdades sociales, raciales y de género son de carácter estructural, en lugar de actos esporádicos meramente mezquinos o irracionales (Young, 2002).

Por su parte, la igualdad clásica, entendida como no discriminación, tiene un carácter neutro, no toca las estructuras de injusticia social y mantiene el *statu quo*. Teniendo presente nuestro contexto histórico es preciso entender el principio de igualdad no solo desde una visión colectiva, sino como un principio antisubordinador (Siegel, 2003; Fiss, 2004). Aquello significa que el propósito del principio de igualdad va más allá de una garantía contra la discriminación, ya que su finalidad es desmantelar las estructuras sociales que discriminan y excluyen a determinados grupos, lo cual está en línea con el sentido dado recientemente por la Comisión Interamericana de Derechos Humanos. Esta perspectiva de igualdad es significativa a la hora de evaluar discriminación en sistemas algorítmicos, ya que clarifica el hecho de que cualquier trato desigual sobre dichos grupos, mediando o no intención e independiente de análisis evaluativos, no es tolerado y, por lo tanto, debe ser sancionado, puesto que su resultado reproduce y perpetúa las condiciones estructurantes de injusticia social.

La discriminación desde una perspectiva colectiva basada en un valor moral

Siguiendo con el punto anterior, la discriminación debe ser entendida, asimismo, desde una visión colectiva, puesto que lo que se afecta es una característica compartida por un grupo de personas, es decir, a todo un conjunto de forma directa al perpetuar y reforzar discriminaciones de carácter estructural.

Asimismo, es menester una concepción de discriminación que pueda hacer frente a nuevas configuraciones, como la discriminación algorítmica, teniendo presente el contexto social pertinente y su evolución. En ese sentido, Issa Kohler Hausmann ha conceptualizado la discriminación como «una acción o práctica que actúa o reproduce un aspecto de una categoría de manera moralmente objetable» (Kohler-Hausmann, 2019).

Según la autora, este concepto tiene un contenido ético complejo,³⁵ en consecuen-

35. La autora hace mención del filósofo moral Bernard Williams, que acuñó la frase «concepto ético grueso» para aquellos términos que describen y evalúan simultáneamente el objeto al que se aplican. Para más detalle véase Williams (1988).

cia, para determinar si un acto ha sido discriminatorio se debe partir de una norma moral, configurando una explicación constitutiva que describe cómo una acción o práctica es moralmente objetable teniendo en consideración el complejo tramado de significados y relaciones sociales que constituyen la categoría social y su contenido estructural (Kohler-Hausmann, 2019). De esta manera, todas las aprensiones sobre cómo podría responder la legislación frente a nuevos grupos, o formas de discriminar a determinados grupos, quedarían contempladas en esta hipótesis reflexiva.

Esta teoría es crítica de la clásica forma de detectar una discriminación, el modelo causal contrafactual, que se centra en una manipulación de una condición protegida, aislándola de su contexto. En ese sentido, frente a una discriminación por causa de género, la causa «mujer» no puede ser considerada como un hecho aislado, sino que debe ser entendida como una construcción producto de una discriminación histórica, sin ser posible simplificarla (Hu y Kohler-Hausmann, 2020).

Lo anterior es importante a la hora de determinar cuándo se produce una discriminación mediante algoritmos, ya que no cualquier modelo de ML, y en particular aquél basado en sistemas de equidad que permiten una prueba estadística, es capaz de reflejar estas construcciones sociales. Por lo tanto, estos métodos cuantitativos necesariamente deben ser complementados con otros de carácter cualitativo.

Aportes de la teoría situada al análisis y fortalecimiento de la legislación: el caso de Chile

En Chile, al igual que en el resto de América Latina, la igualdad formal se encuentra protegida a nivel constitucional, específicamente en los artículos 1, 2 y 19 de la Constitución Política de la República de Chile. Esta clásica definición de igualdad desde la perspectiva personal o individual comprende: 1) la igualdad ante la ley, referida a la igualdad de trato en situaciones similares, y 2) la igualdad en la ley, que es la expresión sustantiva del derecho, igualdad entendida como no discriminación (Nogueira Alcalá, 2006).

Respecto de normas de rango legal, en el año 2012, con el objeto de establecer una normativa especial que reemplazara la función del recurso de protección constitucional por discriminación arbitraria, debido a su poca efectividad y su naturaleza tutelar, que no es compatible con el conocimiento de juicios complejos, entró en vigor la Ley 20.609 que Establece Medidas contra la Discriminación. Esta ley se complementa con la normativa en materia laboral de no discriminación, que posee actualmente una protección de carácter constitucional, denominada «tutela laboral» que, según ha establecido la jurisprudencia, es de amplio espectro, definiéndose que esta discriminación corresponde a cualquier exclusión o menoscabo en el trabajo fundado en criterios distintos de la capacidad o idoneidad del trabajador (Rojas Miño, 2020).

Este concepto bastante amplio se adecúa a la interpretación de la definición de discriminación establecida en el punto anterior.

Por su parte, la referida ley sobre no discriminación arbitraria, define este concepto como toda distinción, exclusión o restricción que carezca de justificación razonable y que cause privación, perturbación o amenaza en el ejercicio legítimo de los derechos fundamentales establecidos en la Constitución Política de la República o en los tratados internacionales sobre derechos humanos. Luego señala que se consideraran razonables las distinciones, exclusiones o restricciones que se encuentren justificadas en el ejercicio legítimo de otro derecho fundamental.

Teniendo presente esta normativa, se puede concluir que actualmente no existe en Chile una protección adecuada respecto de la gran mayoría de los casos de discriminación en base a sistemas automatizados de tomas de decisiones basados en IA, ya que derechamente esta norma no contempla la hipótesis de discriminación indirecta. En ese mismo sentido, denotando su carácter individual, una acción de no discriminación en virtud de esta ley, solo puede ser interpuesta por una sola víctima, contraria a la común ocurrencia, donde las víctimas son generalmente un grupo de personas.

En ese sentido, esta ley presenta profundas falencias. Posee un contenido netamente procesal, estableciéndose en su artículo 1 que su objeto es la instauración de un mecanismo judicial frente a un acto de discriminación arbitraria, en circunstancias en que por su naturaleza esta materia es de contenido sustantivo.

Adicionalmente, esta norma posee disposiciones que desincentivan el ejercicio de la acción, propiamente tal, como la falta de contemplación de indemnización reparatoria de perjuicios, y, por el contrario, se dispone una indemnización imputada a aquel demandante que obtenga un fallo adverso. De la misma manera, no contempla ninguna regla sobre la carga probatoria, por lo que se debe aplicar las normas generales del Código Civil, que deja la carga probatoria a quien presenta la acción. Todo lo anterior genera un escenario complejo y desventajoso para las víctimas de discriminaciones de cualquier naturaleza, y por sobre todo en materias de IA debido a la dificultad adicional de acceder a información relevante respecto de sistemas que por su naturaleza son opacos. En consecuencia, frente a una discriminación a través de un sistema de IA, al menos fuera del ámbito laboral donde podría ser discutible, existe actualmente en Chile una total desprotección legal que no es proporcional al creciente exponencial del uso de este tipo de sistemas, específicamente en grupos vulnerables.

Sin perjuicio de lo anterior, existe actualmente un proyecto de ley que modifica y fortalece la Ley 20.609 (Boletín 12.748-17), el cual está actualmente en trámite legislativo. Este proyecto de ley además de fortalecer en términos generales la protección de las personas frente a actos discriminatorios, protege asimismo contra discriminaciones algorítmicas. Sobre este carácter, se hacen presente las siguientes consideraciones:

- Dentro del proyecto es posible encontrar la dimensión antisubordinadora del principio de igualdad y no solo consagrada como garantía de no discriminación. En efecto, el proyecto modifica el objetivo de la norma indicando que «tiene por finalidad la erradicación, prevención, sanción y reparación de toda discriminación en contra de cualquier persona o grupo de personas». En consecuencia, frente a una discriminación algorítmica, el mero hecho de reproducir una injusticia, aun cuando no fuese intencionada, se castigará, ya que este acto promueve y perpetúa las estructuras sociales que el principio de igualdad busca dismantelar. Sin embargo, considerando que la razón de la subordinación es dismantelar discriminaciones estructurales, se hace necesario, además, la creación de una institución independiente, con autonomía, que cumpla funciones fiscalizadoras, preventivas y educativas para lograr sus fines (artículo 1).
- Definición de discriminación arbitraria: el nuevo proyecto de ley entiende por discriminación arbitraria toda distinción, exclusión, preferencia o restricción carente de justificación razonable, que tenga por objetivo o resultado la perturbación o amenaza en el ejercicio legítimo de los derechos fundamentales. Dichas distinciones, preferencias o restricciones «merecerán particular reproche» cuando se funden en un listado no taxativo de características sospechosas. De esta definición, la expresión «merecerán particular reproche» se sugiere ser eliminada en la propuesta final, toda vez que puede prestarse para interpretar que la norma ampara a otros grupos distintos de aquellos protegidos por encontrarse en una situación de vulnerabilidad, lo cual es contrario al derecho internacional y la generalidad del derecho comparado, ya que la esencia de una norma de no discriminación es la protección exclusiva de grupos en desventaja (artículo 2).
- El proyecto de ley establece y define una discriminación indirecta como aquella que se produce, en la esfera pública o privada, cuando una disposición, un criterio o una práctica aparentemente neutra, es susceptible de implicar una desventaja particular para las personas que pertenecen a un grupo específico. También define la discriminación estructural, que es aquella de carácter grave que se ejerce en perjuicio de grupos vulnerables por su condición, situación social, económica o cultural, quienes han sido históricamente marginados, excluidos o discriminados en sus circunstancias materiales. Si bien lo anterior está en línea con la definición dada en el capítulo previo, tanto de igualdad como de discriminación, en el sentido de no centrar estos conceptos en un elemento comparativo, sino entendiendo las categorías como construcciones sociales creadas por injusticias estructurales, una descripción con todos estos elementos incluidos dentro de una sola clasificación de discriminación directa

e indirecta hubiese sido suficiente y más clara para la ciudadanía (artículo 2b). Hace referencia a un mecanismo similar de «justificación de razonabilidad europeo», que comprende finalidad legítima constitucional, determinar que los medios para alcanzar dicha finalidad sean necesarios y que no generen estigmatización o menoscabo. En este caso se deben tomar los mismos resguardos que en el derecho europeo, en el sentido que las métricas y parámetros evaluativos norteamericanos pueden no encajar bien en un sistema de razonabilidad subjetiva, teniendo en consideración, además, las particularidades de grupos protegidos en Latinoamérica, que como se ha mencionado pueden determinar marcadores y *proxies* racistas no codificados por los sistemas de medición de sesgo algorítmico diseñados en países como Estados Unidos.

- Establece una pluralidad de víctima, también el ejercicio de acciones de interés colectivo o difuso de uno o más grupos o colectivos discriminados; la acción también podrá ser presentada por personas jurídicas sin fines de lucro, constituida a lo menos con tres meses de anterioridad a la presentación de la acción (artículo 4).
- Es importante definir claramente quiénes pueden representar a las víctimas en un proceso por actos discriminatorios, además de su representante legal, tanto de forma individual como colectiva, por la complejidad del tema. En particular, en materias de IA deben ser limitados a agrupaciones idóneas y con conocimiento de esta temática. De lo contrario las víctimas podrían recibir una asesoría deficiente, por falta de experiencia, o ser revictimizadas, esta vez debido a abusos por parte de personas inescrupulosas.
- La regla general es que cuando hablamos de actos discriminatorios efectuados por la acción de un sistema de IA no existe intencionalidad.³⁶ Esto es una diferencia relevante con los actos discriminatorios hechos por personas. Las personas discriminan con intencionalidad negativa, por ejemplo, por racismo o prejuicio, pero en el caso de la IA operan muchas veces sesgos de forma incrustada, lo que pudiese parecer una discriminación más abstracta y por lo tanto difícil de probar (Wachter, Mittelstadt y Russell, 2020) Mittelstadt, y Russell 2020. Este principio debería quedar expresamente establecido en la ley nacional del ramo.
- En ese sentido, es de toda lógica que la evaluación se determine de acuerdo con una norma moral y un análisis contextual social, lo cual desde el punto de vista de la discriminación algorítmica permite llenar vacíos en las actuales

36. Streaming Tool podría constituir una excepción donde se podría alegar una intencionalidad de los funcionarios del Gobierno.

normas del derecho comparado, en complemento con las pruebas estadísticas de acuerdo a la interpretación de discriminación descrita en el acápite anterior.

- No existe mención a pruebas estadísticas en el proyecto de ley quedando como una responsabilidad de los tribunales determinar a través de la jurisprudencia los parámetros evaluativos de esta prueba, para lo cual cobran relevancia las características particulares de la región latinoamericana, la identificación de características interseccionales y *proxies* o marcadores raciales determinados. Deberían ser reemplazados por un enfoque más coherente que contenga modelos cuantitativos, cualitativos y relacionales.
- El proyecto de ley contempla la configuración de discriminación de *prima facie* si a partir de los antecedentes aportados al momento de ingresar la acción, el juez estima la existencia de fundamento plausible de que se ha producido una acción u omisión que importe una diferencia de trato, por lo que así lo señalará al momento de recibir la causa a prueba y corresponderá al denunciado explicar los fundamentos de las medidas adoptadas y de su proporcionalidad (artículo 10), así como un mecanismo indemnizatorio (artículo 14).
- Finalmente, en vista al carácter antisubordinador del principio de igualdad, se hace necesario contar con una institucionalidad autónoma para desarrollar dicho carácter, que cumpla funciones fiscalizadoras y educativas, incorporando auditorías y evaluaciones de impactos algorítmicos. Es así como en Canadá (McKelvey y MacDonald, 2019), por ejemplo, existe una normativa especial e independiente de regulación de sistemas de toma de decisiones automatizadas que incorporan auditorías, registros y evaluaciones de impacto algorítmicas. La evaluación de impacto es de gran relevancia en temas de IA, ya que es posible corregir y prevenir consecuencias negativas y documentar y registrar detalles de proceso que pueden ser utilizados como información estratégica frente a un caso efectivo de discriminación.

Conclusiones

A través del análisis sobre las particularidades técnicas y legales de una problemática compleja y la propuesta de una teoría conceptual sobre discriminación mediante sistemas de IA de tomas de decisiones automatizadas desde la perspectiva latinoamericana, es posible contar con una clara perspectiva sobre cómo funciona nuestra sociedad. En ese sentido, la IA tiene la cualidad paradójica de descubrir la existencia de estructuras de injusticias creadas a lo largo de la historia y de discriminaciones igualmente estructurales que afectan injustamente a un gran número de personas.

En vista de lo anterior y el hecho de que, como se ha comprobado, estamos frente a una temática contextual y social, se hace necesario el desarrollo de investigacio-

nes y propuestas técnicas y regulatorias sobre el análisis de sesgos algorítmicos en Latinoamérica que no solo tengan en consideración particularidades propias de la región, sino que, tomando en cuenta el principio relacional, se propongan como objetivo principal el beneficio comunitario, en particular, el cuidado de aquellos más vulnerables.

En términos particulares, en materia normativa vinculante chilena, se hace indispensable contar con una protección legal contra discriminaciones algorítmicas que sea parte de un marco regulatorio mayor, compuesto por normas existentes y propuestas futuras que tengan por finalidad regular de forma efectiva sistemas de IA que impacten directamente a las personas.

Referencias

- ALPAYDIN, Ethem (2004). *Introduction to Machine Learning*. Cambridge: MIT Press.
- BUOLAMWINI, Joy y Timnit Gebru (2018). «Gender shades: Intersectional accuracy disparities in commercial gender classification». En *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR. Disponible en <https://bit.ly/3wPj5ta>.
- BALDWIN, Robert, Martin Cave y Martin Lodge (2010). *The Oxford handbook of regulation*. Oxford: Oxford University Press.
- BALDWIN, Robert y Martin Cave (2012). *Understanding Regulation: Theory, Strategy, and Practice*. Oxford: Oxford University Press.
- BAROCAS, Solon y Andrew D. Selbst (2016). «Big Data's Disparate Impact». *California Law Review*, 104: 671-732. DOI: [10.2139/ssrn.2477899](https://doi.org/10.2139/ssrn.2477899).
- BENJAMIN, Ruha (2020). «Race After Technology: Abolitionist Tools for the New Jim Code». *Social Forces*, 98 (4): 1-3. DOI: [10.1093/sf/soz162](https://doi.org/10.1093/sf/soz162).
- BIRHANE, Abeba (2020). «Algorithmic Colonization of Africa». *SCRIPTed*, 17 (2): 389-409. DOI: [10.2966/scrip.170220.389](https://doi.org/10.2966/scrip.170220.389).
- . (2021). «Algorithmic Injustice: A Relational Ethics Approach». *Patterns*, 2 (2): 100-205. DOI: [10.1016/j.patter.2021.100205](https://doi.org/10.1016/j.patter.2021.100205).
- BONILLA-SILVA, Eduardo y David G. Embrick (2006). «Racism without racists: "Killing me softly" with color blindness». En *Reinventing critical pedagogy* (pp. 21-34). Lanham: Rowman & Littlefield.
- BROUSSARD, Meredith (2018). *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge: MIT Press.
- BROWN, Anna, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin y Rhema Vaithianathan (2019). «Toward algorithmic accountability in public services: A qualitative study of affected community perspectives on algorithmic decision-making in child welfare services». En *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-12). DOI: [10.1145/3290605.3300271](https://doi.org/10.1145/3290605.3300271).

- COSTANZA-CHOCK, Sasha (2020). *Design Justice: Community-Led Practices to Build the Worlds We Need*. Cambridge: MIT Press. Disponible en <https://bit.ly/3gk6g2R>.
- CREEL, Kathleen y Deborah Hellman (2021). «The Algorithmic Leviathan: Arbitrariness, Fairness, and Opportunity in Algorithmic Decision Making Systems». *Virginia Public Law and Legal Theory Research Paper*, 2021-13. Disponible en <https://bit.ly/2SesjzT>.
- DE BÚRCA, Gráinne (2012). «The Trajectories of European and American Antidiscrimination Law». *The American Journal of Comparative Law*, 60 (1): 1-22. Disponible en <https://bit.ly/2ThaxMJ>.
- DÍAZ DE VALDÉS, J. y José Manuel (2015). «¿Qué clase de igualdad reconoce el Tribunal Constitucional?». *Ius et praxis*, 21 (2): 317-372. DOI: [10.4067/S0718-00122015000200009](https://doi.org/10.4067/S0718-00122015000200009).
- D'IGNAZIO, Catherine y Lauren F. Klein (2020). *Data Feminism*. Cambridge: MIT Press.
- ESCOBAR, Arturo (2019). *Autonomía y diseño: la realización de lo comunal*. Editorial Universidad del Cauca.
- ESPINOSA, Agustín y Rosa María Cueto (2014). «Estereotipos raciales, racismo y discriminación en América Latina». En Elena Zubieta, José Valencia y Gisela Delfino (coordinadores), *Psicología Social y Política: Procesos teóricos y estudios aplicados* (pp. 431-442). Buenos Aires: Eudeba. Disponible en <https://bit.ly/3zAkSnD>.
- ESTERMANN, Josef (2008). *Si el sur fuera el norte: Chakanas interculturales entre Andes y Occidente*. Quito: Abya Yala.
- EUBANKS, Virginia (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. Nueva York: St. Martin's Press.
- FARKAS, Lilla (2011). *How to present a discrimination claim: handbook on seeking remedies under the EU non-discrimination directives*. Luxemburgo: Publications Office of the European Union. Disponible en <https://bit.ly/3cXs6rV>.
- FISS, Owne (2004). «Another equality». *Issues in Legal Scholarship*, 2 (1). DOI: [10.2202/1539-8323.1051](https://doi.org/10.2202/1539-8323.1051)
- FREIRE, Paulo (1970). *Pedagogía del oprimido*. Ciudad de México: Siglo XXI editores.
- FRIEDLER, Sorelle A., Carlos Scheidegger, Suresh Venkatasubramanian, Sonam Choudhary, Evan P. Hamilton y Derek Roth (2019). «A comparative study of fairness-enhancing interventions in machine learning». En *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 329-338). Nueva York: Association for Computing Machinery. DOI: [10.1145/3287560.3287589](https://doi.org/10.1145/3287560.3287589).
- GILMAN, Michele E. (2020). «Poverty Lawgorithms: A Poverty Lawyer's Guide to Fighting Automated Decision-Making Harms on Low-Income Communities». *Data & Society*. Disponible en <https://bit.ly/354hJOH>.
- GIUFFRIDA, Iria (2019). «Liability for AI Decision-Making: Some Legal and Ethical Considerations». *Fordham Law Review*, 88: 439.

- GOODFELLOW, Ian, Yoshua Bengio y Aaron Courville (2016). *Deep Learning*. Cambridge: MIT Press. Disponible en <https://bit.ly/3cWnNx9>.
- GOODMAN, Bryce y Seth Flaxman (2017). «European Union regulations on algorithmic decision-making and a “right to explanation”». *AI magazine*, 38 (3): 50-57. DOI: [10.1609/aimag.v38i3.2741](https://doi.org/10.1609/aimag.v38i3.2741).
- GREEN, Ben y Lily Hu (2018). «The Myth in the Methodology: Towards a Recontextualization of Fairness in Machine Learning». En *Proceedings of the machine learning: the debates workshop*. Disponible en <https://bit.ly/3xCzEZO>.
- GUPTA, Itisha y Garima Nagpal (2020). *Artificial Intelligence and Expert Systems*. Dulles: Mercury Learning and Information.
- HAGERTY, Alexa y Igor Rubinov (2019). «Global AI ethics: a review of the social impacts and ethical implications of artificial intelligence». *arXiv preprint arXiv:1907.07892*.
- HANNA, Alex, Emily Denton, Andrew Smart y Jamila Smith-Loud (2020). «Towards a critical race methodology in algorithmic fairness». En *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 501-512). DOI: [10.1145/3351095.3372826](https://doi.org/10.1145/3351095.3372826).
- HARAWAY, Donna (2013). *Simians, cyborgs, and women: The reinvention of nature*. Nueva York: Routledge.
- HELLMAN, Deborah (2020a). «Measuring algorithmic fairness». *Virginia Law Review*, 106 (4): 811. Disponible en <https://bit.ly/3vJOyva>.
- HELLMAN, Deborah. (2020b). «Sex, Causation, and Algorithms: How Equal Protection Prohibits Compounding Prior Injustice». *Washington University Law Review*, 98 (2): 481-523. Disponible en <https://bit.ly/2SNpR3G>.
- HELLMAN, Deborah y Sophia Moreau (2013). *Philosophical foundations of discrimination law*. Oxford: Oxford University Press.
- HILL COLLINS, Patricia (2002). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. Nueva York: Routledge.
- HU, Lily y Issa Kohler-Hausmann (2020). «What’s Sex Got To Do With Machine Learning». *arXiv preprint arXiv:2006.01770*.
- KOHLER-HAUSMANN, Issa (2019). «Eddie Murphy and the Dangers of Counterfactual Causal Thinking About Detecting Racial Discrimination». *Northwestern University Law Review*, 113 (5): 1163-1228. Disponible en <https://bit.ly/2TTyXw2>.
- IWANICZ-DROZDOWSKA, Małgorzata y Bartosz Witkowski (2016). «Credit Growth in Central, Eastern, and South-Eastern Europe: The Case of Foreign Bank Subsidiaries». *International Review of Financial Analysis*, 43: 146-58. DOI: [10.1016/j.irfa.2015.11.010](https://doi.org/10.1016/j.irfa.2015.11.010).
- KABER, David (2018). «A conceptual framework of autonomous and automated agents». *Theoretical Issues in Ergonomics Science*, 19 (4): 406-430. DOI: [10.1080/1463922X.2017.1363314](https://doi.org/10.1080/1463922X.2017.1363314).

- KAUSHAL, Amit, Russ Altman y Curt Langlotz (2020). «Geographic distribution of US cohorts used to train deep learning algorithms». *Jama*, 324 (12): 1212-1213. DOI: [10.1001/jama.2020.12067](https://doi.org/10.1001/jama.2020.12067).
- KHAITAN, Tarunabh (2015). *A theory of discrimination law*. Oxford: Oxford University Press.
- KITCHING, Kevin (2005). *Non-discrimination in international law: A handbook for practitioners*. Londres: Interights. Disponible en <https://bit.ly/3xvW3Hm>.
- LANDER, Edgardo y Santiago Castro-Gómez (editores) (2000). *La colonialidad del saber: eurocentrismo y ciencias sociales: perspectivas latinoamericanas*. Buenos Aires: Consejo Latinoamericano de Ciencias Sociales-CLACSO. Disponible en <https://bit.ly/2SjSp4I>.
- LATONERO, Mark (2018). «Governing artificial intelligence: Upholding human rights & dignity». *Data & Society*: 1-37. Disponible en <https://bit.ly/3qbxJZ7>.
- LEHR, David y Paul Ohm (2017). «Playing with the data: what legal scholars should learn about machine learning». *UCDL Rev.*, 51: 653. Disponible en <https://bit.ly/3qblgVn>.
- LIPTON, Zachary C. (2018). «The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery». *Queue*, 16 (3): 31-57. DOI: [10.1145/3236386.3241340](https://doi.org/10.1145/3236386.3241340).
- LIZCANO FERNÁNDEZ, FRANCISCO (2005). «Composición étnica de las tres áreas culturales del continente Americano al comienzo del siglo XXI». *Convergencia*, 12 (38): 185-232. Disponible en <https://bit.ly/2UcpqjG>.
- MALIK, Momin M. (2020). «A Hierarchy of Limitations in Machine Learning». *arXiv preprint arXiv:2002.05193*. Disponible en <https://bit.ly/3zp1p9t>.
- MCKELVEY, Fenwick y Margaret MacDonald (2019). «Artificial intelligence policy innovations at the Canadian federal government». *Canadian Journal of Communication*, 44 (2): 43-50. DOI:[10.22230/cjc.2019v44n2a3509](https://doi.org/10.22230/cjc.2019v44n2a3509)
- MEHRABI, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman y Aram Galstyan
—. (2019). «A Survey on Bias and Fairness in Machine Learning». Disponible en <https://bit.ly/3cxOGqX>.
- MENDEZ, Xhercis (2015). «Notes toward a decolonial feminist methodology: revisiting the race/gender matrix». *Trans-scripts*, 5: 41-56. Disponible en <https://bit.ly/3qdcRjY>.
- MEZA SALCEDO, Guillermo (2017). «Ética de la investigación desde el pensamiento indígena: derechos colectivos y el principio de la comunalidad». *Revista de Bioética y Derecho*, 41: 141-59. Disponible en <https://bit.ly/3vIoiS3>.
- MHLAMBI, Sabelo (2020). «From rationality to relationality: ubuntu as an ethical and human rights framework for artificial intelligence governance». *Carr Center for Human Rights Policy Discussion Paper Series*, 9. Disponible en <https://bit.ly/3qdcRjY>.

ly/3xyqdtK.

- MIGNOLO, Walter D. (2003). *Historias locales/diseños globales: colonialidad, conocimientos subalternos y pensamiento fronterizo*, vol. 18. Madrid: Ediciones Akal.
- MORENO FIGUEROA, Mónica G. y Emiko Saldívar Tanaka (2016). «“We Are Not Racists, We Are Mexicans”: Privilege, Nationalism and Post-Race Ideology in Mexico». *Critical Sociology*, 42 (4-5): 515-533. DOI: [10.1177/0896920515591296](https://doi.org/10.1177/0896920515591296).
- NOBLE, Safiya Umoja (2018). *Algorithms of oppression: How search engines reinforce racism*. Nueva York: New York University Press.
- ORWAT, Carsten (2020). *Risks of Discrimination through the Use of Algorithms*. Berlín: Federal Anti-Discrimination Agency. Disponible en <https://bit.ly/3iS1VH3>.
- PAGE, John, Michael Bain y Faqihza Mukhlis (2018). «The risks of low level narrow artificial intelligence». En 2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR) (pp. 1-6). IEEE. DOI: [10.1109/IISR.2018.8535903](https://doi.org/10.1109/IISR.2018.8535903).
- PASQUALE, Frank (2017). «Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society». *Ohio State Law Journal*, 78: 1243-1255. Disponible en <https://bit.ly/3pKunvR>.
- PEDREGOSA, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay (2011). «Scikit-learn: Machine Learning in Python». *The Journal of machine Learning research*, 12: 2825-2830. Disponible en <https://bit.ly/3cXmQVd>.
- QUIJANO, Aníbal (2000). «Colonialidad del poder, eurocentrismo y América Latina». Buenos Aires: Consejo Latinoamericano de Ciencias Sociales (CLACSO). Disponible en <https://bit.ly/2TQi8BW>.
- RAWLS, John (1990). «Justicia como “fairness”: política, no metafísica». *Revista de Ciencia Política*, 12 (1-2): 89-118. Disponible en <https://bit.ly/3pAkAZi>.
- ROJAS MIÑO, Irene (2020). «Los instrumentos jurídicos ante la discriminación remunerativa entre mujeres y hombres en Chile». *Revista de Derecho* (Universidad Austral de Chile), 33 (2): 145-165. DOI: [10.4067/S0718-09502020000200145](https://doi.org/10.4067/S0718-09502020000200145).
- RONCONI, Liliana (2018) «Repensando el principio de igualdad: repensando el principio de igualdad». *Isonomía: Revista de teoría y filosofía del derecho*, 49: 103-140. DOI: [10.5347/49.2018.25](https://doi.org/10.5347/49.2018.25).
- SALGADO, Mauricio y Javier Castillo (2018). «Differential Status Evaluations and Racial Bias in the Chilean Segregated School System». *Sociological Forum*, 33 (2): 354-377. DOI: [10.1111/socf.12426](https://doi.org/10.1111/socf.12426).
- SAMBASIVAN, Nithya, Erin Arnesen, Ben Hutchinson, Tulsee Doshi y Vinodkumar Prabhakaran (2021). «Re-imagining Algorithmic Fairness in India and Beyond». *arXiv preprint arXiv:2101.09995*.
- SÁNCHEZ-MONEDERO, Javier, Lina Dencik y Lilian Edwards (2020). «What does it

- mean to solve the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems». En *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 458-468). Nueva York: Association for Computing Machinery. arXiv:1910.06144v2.
- SCHNEEBERGER, David, Karl Stöger y Andreas Holzinger (2020). «The European Legal Framework for Medical AI». En Andreas Holzinger, Peter Kiesenberg, A Min Tjoa y Edgar Weippl (editores), *Machine Learning and Knowledge Extraction* (pp. 209-226). Cham: Springer International Publishing. DOI: [10.1007/978-3-030-57321-8_12](https://doi.org/10.1007/978-3-030-57321-8_12).
- SCHUETT, Jonas (2019). «A Legal Definition of AI». DOI: [10.2139/ssrn.3453632](https://doi.org/10.2139/ssrn.3453632).
- SELBST, Andrew D. (2017). «Disparate Impact in Big Data Policing». *Georgia Law Review*, 52: 109-195. DOI: [10.2139/ssrn.2819182](https://doi.org/10.2139/ssrn.2819182).
- SIEGEL, Reva B. (2003). «Equality talk: Antisubordination and anticlassification values in constitutional struggles over Brown». *Harvard Law Review*, 117: 1470-1547. Disponible en <https://bit.ly/3qkgmVZ>.
- SILVA, Graziella Moraes, Emiko Saldivar, Graziella Moraes Silva y Emiko Saldivar (2018). «Comparing Ideologies of Racial Mixing in Latin America: Brazil and Mexico». *Sociologia & Antropologia*, 8 (2): 427-456. DOI: [10.1590/2238-38752017v824](https://doi.org/10.1590/2238-38752017v824).
- SURESH, Harini y John V. Gutttag (2019). «A framework for understanding unintended consequences of machine learning». *arXiv preprint arXiv:1901.10002*.
- . (2020). «A Framework for Understanding Unintended Consequences of Machine Learning». *arXiv:1901.10002 [cs, stat]*, febrero. Disponible en <https://bit.ly/2TUHizz>.
- TUBARO, Paola y Antonio A. Casilli (2019). «Micro-Work, Artificial Intelligence and the Automotive Industry». *Journal of Industrial and Business Economics*, 46 (3): 333-345. DOI: [10.1007/s40812-019-00121-1](https://doi.org/10.1007/s40812-019-00121-1).
- TVERSKY, Amos, y Daniel Kahneman (1974). «Judgment under uncertainty: Heuristics and biases». *science*, 185 (4157): 1124-1131. DOI: [10.1126/science.185.4157.1124](https://doi.org/10.1126/science.185.4157.1124).
- VARELA, Francisco J, Evan Thompson y Eleanor Rosch (2017). *The Embodied Mind, revised edition: Cognitive Science and Human Experience*. Cambridge: MIT Press.
- VENEKLASSEN, Lisa, Valerie Miller, Debbie Budlender y Cindy Clark (2002). *A new weave of power, people & politics: the action guide for advocacy and citizen participation*. Oklahoma City: World Neighbors. Disponible en <https://bit.ly/35FyTT6>.
- WACHTER, Sandra (2020). «Affinity profiling and discrimination by association in online behavioural advertising». *Berkeley Technology Law Journal*, 35 (2). DOI: [10.2139/SSRN.3388639](https://doi.org/10.2139/SSRN.3388639).
- WACHTER, Sandra, Brent Mittelstadt y Chris Russell (2020). «Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI». DOI: [10.2139/ssrn.3547922](https://doi.org/10.2139/ssrn.3547922).

- WADE, Peter (2017). «Racism and Race Mixture in Latin America». *Latin American Research Review*, 52 (3): 477-85. DOI: [10.25222/larr.124](https://doi.org/10.25222/larr.124).
- WIEVIORKA, Michel (2002). *El racismo: Una introducción*. La Paz: Plural editores.
- WILLIAMS, Bernard (1988). «Ethical consistency». En Geoffrey Sayre-McCord (editor), *Essays on moral realism* (pp.41-58). Nueva York: Cornell University Press.
- XIANG, Alice y Inioluwa Deborah Raji (2019). «On the Legal Compatibility of Fairness Definitions». *ArXiv:1912.00761 [Cs, Stat]*, noviembre. Disponible en <https://bit.ly/3pPm8Pn>.
- YEUNG, Karen, Andrew Howes y Ganna Pogrebna (2020). «AI Governance by Human Rights–Centered Design, Deliberation, and Oversight». En Markus Dirk Dubber, Frank Pasquale y Sunit Das (editores), *The Oxford Handbook of Ethics of AI* (pp. 77-106). Oxford: Oxford University Press. DOI: [10.2139/ssrn.3435011](https://doi.org/10.2139/ssrn.3435011).
- YOUNG, Iris Marion (2001). «Equality of whom? Social groups and judgments of injustice». *Journal of political philosophy*, 9 (1): 1-18. DOI: [10.1111/1467-9760.00115](https://doi.org/10.1111/1467-9760.00115).
- YOUNG, Iris Marion (2002). *Inclusion and Democracy*. Oxford: Oxford University Press.
- YU, Ai (2020). «Digital surveillance in post-coronavirus China: A feminist view on the price we pay». *Gender, Work, and Organization*, 27 (5): 774-777. DOI: [10.1111/gwao.12471](https://doi.org/10.1111/gwao.12471).
- ZEMEL, Rich, Yu Wu, Kevin Swersky, Toni Pitassi y Cynthia Dwork (2013). «Learning fair representations». En *International conference on machine learning*, (pp. 325-333). PMLR. DOI: [10.5555/3042817.3042973](https://doi.org/10.5555/3042817.3042973).
- ZUBOFF, Shoshana (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Nueva York: Public Affairs.

Sobre la autora

CATHERINE MUÑOZ GUTIÉRREZ es abogada, licenciada en Ciencias Jurídicas y Sociales, Universidad Central de Chile, Máster (LL.M.) en Derecho Internacional, Inversiones, Comercio y Arbitraje, Universidad de Chile. Master of Laws in International Law (LL.M. int.), Universidad de Heidelberg, Alemania. Directora del Observatorio para la Transparencia e Inclusión Algorítmica (OPTIA), miembro de The Global AI Ethics Consortium (GAIEC). Su correo electrónico es cmunoz@idoneaia.cl. 
<https://orcid.org/0000-0003-1659-8572>.

La *Revista de Chilena de Derecho y Tecnología* es una publicación académica semestral del Centro de Estudios en Derecho Informático de la Facultad de Derecho de la Universidad de Chile, que tiene por objeto difundir en la comunidad jurídica los elementos necesarios para analizar y comprender los alcances y efectos que el desarrollo tecnológico y cultural han producido en la sociedad, especialmente su impacto en la ciencia jurídica.

EDITOR GENERAL

Daniel Álvarez Valenzuela
(dalvarez@derecho.uchile.cl)

SITIO WEB

rchdt.uchile.cl

CORREO ELECTRÓNICO

rchdt@derecho.uchile.cl

LICENCIA DE ESTE ARTÍCULO

Creative Commons Atribución Compartir Igual 4.0 Internacional



La edición de textos, el diseño editorial
y la conversión a formatos electrónicos de este artículo
estuvieron a cargo de Tipografía
(www.tipografica.io).