



UNIVERSIDAD DE CHILE
Facultad de Economía y Negocios

SESGOS ALGORITMICOS, ¿EL NEURODERECHO FALTANTE EN LA LEY CHILENA?

AFE para optar al grado de
Magíster en Políticas Públicas

Alumno: José Fernández López

Profesor Guía: Andrés Gómez-Lobo Echenique

Santiago, agosto de 2022

RESUMEN

En las últimas décadas, la utilización de algoritmos para la toma de decisiones ha crecido exponencialmente. Esto se ha visto facilitado por el aumento en la cantidad de datos disponibles, así como el avance de tecnologías de procesamiento de datos. Una amenaza potencialmente presente en los algoritmos de toma de decisiones es la de sesgo algorítmico, entendido como la tendencia a discriminar arbitrariamente ciertos grupos sociales (según género, etnia, pensamiento político, entre otros). Para resguardar a los afectados ante las situaciones anteriores, puede ser conveniente adaptar la legislación y/o institucionalidad nacional, mediante la incorporación de medidas que permitan tratar esta amenaza. Este trabajo discute sobre la necesidad de incorporar el derecho a la protección contra los sesgos algorítmicos en la legislación. Paralelamente, planteamos de manera exploratoria aspectos a considerar en una institucionalidad que facilite dicha protección. El concepto de transparencia algorítmica parece ser central para lograr aquello.

Palabras clave: neuroderechos, algoritmos, sesgos algorítmicos, discriminación, transparencia algorítmica

I. INTRODUCCIÓN

Los algoritmos juegan un rol crítico en la mayoría de los sistemas computacionales, particularmente en los sistemas computacionales autónomos. Debido al aumento exponencial en la creación de datos, el uso de sistemas computacionales autónomos para el tratamiento de la información también ha aumentado en las últimas décadas, precisamente porque estos sistemas no tienen algunos de los defectos que tenemos los humanos, en cuanto a capacidad y rapidez en la ejecución de procesos relacionados con datos. Si los sistemas autónomos van a ser las mejores versiones de nosotros (al menos para algunas tareas), entonces deberíamos aspirar a utilizar la mayoría de los algoritmos que podamos (Danks y London, 2017). Sin embargo, pueden darse casos de parcialidades (intencionadas o no) en contra de ciertos grupos sociales en la utilización de algoritmos. Esto es lo que denominaremos preliminarmente como "sesgo algorítmico".

Una de las características de los algoritmos de inteligencia artificial es que, frecuentemente, sufren de opacidad, por lo que es difícil obtener una idea de su mecanismo interno de trabajo, lo que complica aún más el problema, porque confiar decisiones importantes a un sistema que no puede explicarse por sí mismo presenta peligros evidentes (Adadi y Berrada, 2018). Dado lo anterior, es que puede existir la posibilidad de escasa auditabilidad en la elaboración de los algoritmos, por lo que se pueden generar episodios de sesgo en la elaboración o tratamiento de estos.

Dado que los sesgos algorítmicos están incluidos en distintos procesos cotidianos, es posible determinar situaciones en que las personas pueden ser afectadas por estos sesgos, por ejemplo, que un software para predecir futuros criminales este sesgado contra la gente de raza negra (Angwin et al. 2016), o cuando en el año 2014 Amazon intentó desarrollar un algoritmo de reclutamiento, el cual favorecía a los candidatos hombres mientras penalizaba a las mujeres, para lo cual los curriculums que contenían la palabra “de mujeres” o los nombres de universidad sólo para mujeres fueron bajados de categoría, mientras que favorecieron los curriculums que contenían lenguaje masculino asociado con ingenieros varones (Dastin, 2018).

En relación con la reciente aprobación del proyecto de ley chileno que establece los neuroderechos como derechos a ser incluidos dentro de la Constitución, actualmente este proyecto de ley no contempla el neuroderecho a la protección contra sesgos algorítmicos. De todas formas, dicho proyecto de ley es pionero en su materia; en efecto, las potencias mundiales a la fecha carecen de leyes que abarquen de manera explícita este tipo de derechos. Le sigue España, país que en el año 2021 aprueba la Carta de los Derechos Digitales, la cual promueve de manera no vinculante la protección de los neuroderechos, incluyendo aquel referido a la protección contra sesgos algorítmicos. La falta de legislación en torno a los neuroderechos es una de las razones por las cuales las posturas existentes son escasas, sin embargo, la falta de protección ante los sesgos algorítmicos puede tener efectos negativos importantes.

En este documento discutiremos la necesidad, desde un punto de vista de política pública, de incluir en la Constitución el derecho a la protección contra sesgos algorítmicos. Además, proponemos, de manera exploratoria, avances institucionales concretos que pudieran facilitar dicha protección. El resto del documento se organiza como sigue. En la sección II se revisará el origen de los neuroderechos. En la sección III haremos una revisión del concepto de sesgos

algorítmicos. En la sección IV se revisarán los avances en materia legislativa en Chile respecto al tema. En la sección V se analizará una posible propuesta de protección de los neuroderechos, para finalmente concluir en la sección VI.

II ORIGEN DE LOS NEURODERECHOS

Para el año 2028 se estima que habrá 150 billones de sensores midiendo todo lo que hacemos (Jackson, 2018), lo cual generará cantidades gigantescas de información. Debido a esta cantidad de información es que en las últimas décadas el avance de las tecnologías que utilizan la automatización de datos ha puesto en la mira los derechos humanos básicos universales, dado que muchas normativas no contemplan derechos que tengan relación con la interacción del ser humano con estas tecnologías, sobre todo con las vulnerabilidades que puedan ser expuestas producto de la interacción humano-maquina. Dado lo anterior es que se hace necesario tener un nuevo marco legal que regule este tipo de derechos, denominados “neuroderechos”.

El origen de los neuroderechos, como concepto, se establece en el año 2017 a partir de la NeuroRights Initiative, perteneciente a la NeuroRights Foundation¹. Dicha iniciativa promueve la protección de 5 neuroderechos: 1) el derecho a la identidad personal; 2) el derecho al libre albedrío; 3) el derecho a la privacidad mental; 4) el derecho al acceso equitativo a potenciadores mentales; 5) el derecho a la protección contra sesgos algorítmicos (Yuste et al., 2021). A continuación, señalamos a qué se refiere con cada uno de ellos:

- 1) El derecho a la identidad personal: deben desarrollarse límites para prohibir que la tecnología altere el concepto del uno mismo. Al conectar el cerebro de individuos a computadoras, la Neurotecnología² podría nublar la línea entre la conciencia de una persona y las tecnologías externas.
- 2) El derecho al libre albedrío: Las personas deben tener control absoluto sobre sus propias decisiones, sin la manipulación de Neurotecnologías externas.
- 3) El derecho a la privacidad mental: todos los datos obtenidos al medir la actividad neuronal ("NeuroDatos") deben mantenerse privados. Adicionalmente, la venta, la transferencia comercial y el uso de datos neuronales deben estar estrictamente regulados.
- 4) El derecho al acceso equitativo a potenciadores mentales: deben establecerse pautas nacionales e internacionales que regulen el desarrollo y las aplicaciones de Neurotecnologías que permitan potenciar las capacidades mentales. Estas directrices deberán sustentarse en el principio de justicia y garantizar la igualdad de acceso a todos los ciudadanos.
- 5) El derecho a la protección contra sesgos algorítmicos: Utilizar contramedidas para combatir sesgos algorítmicos debe ser la norma al emplear algoritmos de aprendizaje automático. Se deben considerar indicaciones de distintos grupos de usuarios para enfrentar los sesgos de manera colectiva desde las primeras fases del desarrollo del algoritmo.

¹ <https://neurorightsfoundation.org/>

² Neurotecnología se define como como el conjunto de métodos e instrumentos que permiten una conexión directa de componentes técnicos con el sistema nervioso (Müller y Rotter, 2017).

Sin embargo, estos son términos muy genéricos, a menudo sujetos a interpretación, para los cuales aún no existe un marco integral para abordar el alcance y la gama más amplia de posibles violaciones de los neuroderechos.

Actualmente, no existe un consenso internacional sobre lo que constituye neuroderechos, dado que es un tema que continúa siendo debatido en la actualidad. Chile es el único país con un proyecto de ley y una modificación constitucional que exigen la neuroprotección y protegen explícitamente los neuroderechos (Yuste et al. 2021). Adicionalmente le sigue España, el cual representa otro esfuerzo pionero en explorar el panorama de los derechos humanos de la era digital, incorporando los 5 neuroderechos propuestos por la Neurorights Foundation, a través de la Carta de Derechos Digitales.

III. SESGOS ALGORÍTMICOS

La cantidad de datos creados por y acerca de nosotros ha crecido exponencialmente. Para el caso de los algoritmos, su finalidad es buscar patrones en los datos, usando datos del pasado e infiriendo significado de datos faltantes, para los cuales estos modelos pueden aprender patrones para correlacionarlos con variables sociales (ej: sexo o raza) y producir resultados discriminatorios, ya sean intencionados o no (Jackson, 2018).

El algoritmo, como procedimiento sistemático y efectivo para resolver un problema, se compone de un número concreto de instrucciones que deben estar bien definidas para operar sobre un tipo concreto de datos, a través de un número finito de pasos, que aporta una solución, generalmente matemática, a todos los casos analizados. La solidez de la formulación de un algoritmo dependerá de que sea sistemático y consistente, pero también de que sea no ambiguo, es decir que no deje abierta la solución a interpretaciones. La misión de un algoritmo no es el cálculo de una solución exacta o precisa, sobre todo porque como objetivo se convierte en inalcanzable en muchos casos; su misión está más próxima a facilitar una respuesta satisfactoria o válida, es decir, útil más que en su neutralidad o imparcialidad (Benítez, 2019).

Los algoritmos juegan un rol clave en el funcionamiento de los sistemas autónomos, por lo que periódicamente se han planteado preocupaciones sobre la posibilidad de un sesgo algorítmico que pueda derivar de la utilización de estos tipos de sistemas. Sin embargo, los debates en esta área se han visto obstaculizados por diferentes significados y usos del término “sesgo algorítmico”. A veces se usa como un término puramente descriptivo, a veces como un término peyorativo, y tales variaciones pueden generar confusión y dificultar las discusiones sobre cuándo y cómo responder al sesgo algorítmico (Danks y London, 2017). En otros trabajos el término “algorítmico” se ha utilizado para referirse a los sistemas automatizados de toma de decisiones en general, con productos y procesos de aprendizaje automático que desempeñan un papel cada vez mayor en dichos sistemas automatizados. En cuanto al término “sesgo”, el diccionario de la Real Academia Española (2022) lo define como “un error sistemático en el que se puede incurrir cuando al hacer muestreos o ensayos se seleccionan o favorecen unas respuestas frente a otras”. Banu et al. (2020) definen el término sesgo algorítmico como “discriminación que es sistémica e injusta al favorecer ciertos grupos individuales sobre otros”. Moya y Le (2021) señalan que el

sesgo algorítmico “ocurre cuando una decisión algorítmica crea resultados injustos que privilegian injustificada y arbitrariamente a ciertos grupos sobre otros”.³

El término sesgo algorítmico se ha aplicado a una variedad de ejemplos de injusticia en los sistemas automatizados, solo algunos de los cuales parecen ajustarse a las definiciones estadísticas o técnicas de sesgo. En respuesta, a medida que han aparecido ejemplos de uso popular, los investigadores han intentado describir los criterios de sesgo. Baker y Hawn (2021) señalan que la mayoría de los ejemplos de sesgo algorítmico se extraen de casos en los que se utilizaron grandes conjuntos de datos para entrenar modelos predictivos utilizando métodos algorítmicos o estadísticos

En cuanto a los tipos de sesgos, Baeza-Yates y Peiró (2019) indican que los sesgos más conocidos son: el sesgo estadístico (relacionado a cómo obtenemos los datos, de errores de medida o similares); el sesgo cultural (aquel que deriva de la sociedad, del lenguaje que hablamos o de todo lo que hemos aprendido a lo largo de la vida, por ejemplo, los estereotipos); y el sesgo cognitivo (es aquel que nos identifica y que depende de nuestra personalidad, de nuestros gustos y miedos). En relación con lo anterior es que los tipos de sesgos relacionados con sesgos algorítmicos corresponderían a sesgos culturales y sesgos cognitivos, principalmente.

El creciente uso de algoritmos en la vida social y económica ha generado una preocupación: que se puedan discriminar inadvertidamente a ciertos grupos. Por ejemplo, un estudio reciente descubrió que los algoritmos de procesamiento del lenguaje natural pueden incorporar sesgos de género básicos, como asociar la palabra enfermera más estrechamente con la palabra “ella” que con la palabra “él”, debido a que los datos utilizados para entrenar estos algoritmos están teñidos de estereotipos y discriminación pasada (Kleinberg et al., 2018). También, ejemplo de lo anterior, es el reciente trabajo de Fuller et al. (2021), en donde se menciona a los trabajadores que son sometidos a procesos automáticos de contratación relacionados con difusión de puestos de trabajo, en donde existen procesos tales como publicidad basada en algoritmos, clasificación de curriculums por algoritmos y evaluación e incorporación automatizadas, lo anterior con un componente de sesgo y posible discriminación en el proceso y que finalmente conllevan a no contratar a los posibles candidatos idóneos para los puestos de trabajos disponibles.

En esta misma línea, existe una preocupación creciente de que los algoritmos puedan generar discriminaciones raciales y de género a través de las personas que los construyen o a través de los datos utilizados para entrenarlos. El trabajo empírico está prestando cada vez más apoyo a estas preocupaciones. Por ejemplo, Datta et al. (2015) encontraron evidencia que Google usó el género para determinar los anuncios a presentar a hombres versus mujeres; a los hombres se le mostró anuncios que los motivaba a buscar servicios de coaching con trabajos mejores pagados, mientras que a las mujeres se les mostró anuncios más genéricos. Los sistemas de reconocimiento facial que se utilizan cada vez más en la aplicación de la ley funcionan peor en

³ La bibliografía revisada hace referencia a la discriminación hacia grupos de personas, no a individuos en particular. También podría existir discriminación positiva, sin embargo, el presente trabajo se centra en la discriminación negativa hacia grupos de personas.

el reconocimiento de rostros de mujeres y personas negras, y los algoritmos de procesamiento del lenguaje natural codifican el lenguaje en formas de género (Obermeyer et al. 2019).

A comienzos del 2014, como respuesta a las crecientes preocupaciones sobre el papel que juegan los algoritmos de minería de datos en la toma de decisiones, el presidente de los Estados Unidos Barack Obama pidió una revisión de las prácticas de recopilación y análisis de *big data*. El informe resultante concluyó que las tecnologías de *big data* pueden causar daños sociales más allá de los daños a la privacidad. En particular, expresó su preocupación por la posibilidad de que las decisiones informadas por *big data* pudieran tener efectos discriminatorios, incluso en ausencia de intención discriminatoria, imponiendo un trato menos favorable a grupos ya desfavorecidos (Hajian et al. 2016).

Bozdog (2013) señala que las personas afectan el diseño de los algoritmos, pero también pueden influir manualmente en el proceso de filtrado después de que el algoritmo ha sido diseñado. Por lo tanto, cambiar de edición humana a control algorítmico no elimina todos los sesgos humanos. Además, los factores individuales, como los juicios personales, los factores organizacionales, como las políticas de la empresa, los factores externos, como las solicitudes del gobierno, seguirán estando presentes debido al papel de los seres humanos en la prestación de estos servicios.

IV. PROYECTO DE LEY CHILENO EN NEURODERECHOS

La ley N° 21.383 de 2021 modifica el artículo 19 de la Constitución chilena, con el objetivo de “consagrar la protección de la integridad y la indemnidad mental en relación con el avance de los neuroderechos” (Boletín N° 13.828-19). La reforma constitucional entiende el concepto de neuroderechos como, primero, el derecho a la privacidad mental (es decir, que por ejemplo los datos del cerebro de las personas se traten con una confidencialidad equiparable al de los trasplantes de órganos) y segundo, como el derecho a la identidad, manteniendo la individualidad de las personas. Adicionalmente, dispone que: “El desarrollo científico y tecnológico estará al servicio de las personas y se llevará a cabo con respeto a la vida y a la integridad física y psíquica. La ley regulará los requisitos, condiciones y restricciones para su utilización en las personas, debiendo resguardar especialmente la actividad cerebral, así como la información proveniente de ella” (Biblioteca del Congreso Nacional, 2021).

Actualmente se encuentra en tramitación en el Congreso un proyecto de ley (Boletín N° 13.828-19) que tiene por objetivo especificar dicha regulación. El proyecto de ley posee un marcado anclaje en el respeto a la dignidad humana como principio subyacente al que deberían someterse las neurotecnologías, incorporando, además, un elemento de igualdad de acceso frente al desarrollo de la técnica, que se materializa a través del igual acceso al aumento de la capacidad mental, para evitar cualquier atisbo de diferenciaciones arbitrarias, e ilícitas (Boletín N° 13.828-19).

Entre los objetivos generales que pretende establecer el proyecto de ley chileno se considera:

“a) proteger la integridad física y psíquica de las personas, a través de la protección de la privacidad de los datos neuronales, del derecho a la autonomía o libertad de decisión individual, y del acceso sin discriminaciones arbitrarias a aquellas neurotecnologías que conlleven aumento de las capacidades psíquicas.

b) fomentar la concordancia entre el desarrollo de neurotecnologías e investigación médico-clínica con los principios éticos de la investigación científica y médica y así sean favorables al bien y beneficio común.

c) garantizar la información a los usuarios de neurotecnologías sobre sus potenciales consecuencias negativas y efectos secundarios, y el derecho al control voluntario sobre el funcionamiento de cualquier dispositivo conectado a su cerebro.” (Boletín N° 13.828-19).

Dado lo anterior, es que el proyecto de ley pretende resguardar 3 de los 5 neuroderechos: el derecho a la privacidad mental, el derecho al libre albedrío y el derecho al acceso equitativo a potenciadores mentales, en concordancia con la ley 21.383 ya mencionada.

El proyecto de ley chileno se enmarca en el debate mundial respecto al uso de herramientas de automatización, que en muchos casos puede generar externalidades que puedan afectar a las personas que se ven sometidas a estos procesos de automatización. Tal es el caso de España y su Carta de Derechos Digitales, la cual, si bien no tiene un carácter normativo, ofrece un marco de referencia para garantizar los derechos de las personas ante los entornos virtuales y digitales. El documento se refiere a la protección contra posibles sesgos algorítmicos específicamente en las siguientes secciones:

“a) Derecho a la igualdad y a la no discriminación en el entorno digital: El derecho y el principio a la igualdad inherente a las personas será aplicable en los entornos digitales, incluyendo la no discriminación y la no exclusión. En particular, se promoverá la igualdad efectiva de mujeres y hombres en entornos digitales. Se fomentará que los procesos de transformación digital apliquen la perspectiva de género adoptando, en su caso, medidas específicas para garantizar la ausencia de sesgos de género en los datos y algoritmos usados.

b) Derechos en el ámbito laboral: La garantía de sus derechos frente al uso por la entidad empleadora de procedimientos de analítica de datos, inteligencia artificial y, en particular, los previstos en la legislación respecto del empleo de decisiones automatizadas en los procesos de selección de personal.

c) Derechos digitales en el empleo de las neurotecnologías: Asegurar que las decisiones y procesos basados en neurotecnologías no sean condicionadas por el suministro de datos, programas o informaciones incompletos, no deseados, desconocidos o sesgados.” (Carta de Derechos Digitales, 2021).

Es posible notar que, respecto a la sección que dispone el derecho a la igualdad y a la no discriminación en el entorno digital, este se ve vulnerado en ejemplos revisados al principio de este trabajo, en donde se evidenciaron situaciones de discriminación racial y por género en entornos digitales.

Y, en relación a los derechos en el ámbito laboral, mencionamos el caso de selección de personal por procesos automatizados que utilizan algoritmos y que favorecen a algunos tipos de candidatos en vez de otros, generando un perjuicio a posibles candidatos que puedan cumplir con los requisitos para el cargo ofertado.

En cuanto a los derechos digitales en el empleo de las neurotecnologías, cabe mencionar que estos también son contemplados como uno de los neuroderechos a ser protegidos en el proyecto de ley chileno (Boletín N° 13.828-19).

V. PROPUESTA DE PROTECCIÓN

Antes de comenzar a analizar posibles propuestas de protección ante sesgos algorítmicos, es necesario determinar el origen del sesgo, que podría clasificarse en un nivel general como sesgo voluntario y sesgo involuntario, de acuerdo con lo revisado en la literatura. El sesgo voluntario, lo definiremos como aquel en el cual los responsables de la elaboración y utilización del algoritmo incluyen características sesgadas, es decir, intencionadamente la elaboración del algoritmo se configura de manera tal que los resultados generen un sesgo deseado, generalmente con la finalidad de discriminar a unas observaciones de otras. El sesgo involuntario corresponde a aquel en el cual no existe una finalidad deseada de discriminar observaciones.

En una primera aproximación se podría concluir que es muy delgada la línea que podría diferenciar un sesgo algorítmico de un error algorítmico. Para resolver esta problemática se hace necesario generar un proceso denominado “transparencia algorítmica”, es decir, la capacidad para saber qué datos se utilizan, cómo se utilizan, quiénes los utilizan, para qué los utilizan y cómo se llega a partir de los datos a tomar las decisiones que afectan a la esfera vital de quien reclama esta transparencia (Araya, 2021). Por tanto, la presente propuestas de protección utiliza los criterios de transparencia algorítmica desarrollados por Diakopoulos (2016), quien señala la posibilidad de generar un estándar de transparencia algorítmica, que tenga foco en las siguientes 5 dimensiones:

1) Participación humana: la transparencia en torno a la participación humana debiera explicar el objetivo del algoritmo, incluido el proceso editorial humano incluido en su elaboración. En cualquier acción colectiva, es difícil desagregar y asignar crédito exactamente a quién hizo qué (o podría ser responsable de un error en particular), no obstante, las personas involucradas pueden sentir un mayor sentido de responsabilidad pública y presión si sus nombres están en juego.

2) Datos: una vía para la transparencia es comunicar la calidad de los datos, incluida su precisión, integridad e incertidumbre, así como su puntualidad (ya que la validez de los datos puede cambiar a través del tiempo), representatividad de una muestra para una población específica, y suposiciones u otras limitaciones. También se pueden transparentar otras dimensiones del procesamiento de datos: ¿cómo se definió, recopiló, transformó, examinó y editó (ya sea automáticamente o por intervención humana) ?, ¿cómo se recopilan varias etiquetas de datos? ¿Reflejan un proceso más objetivo o subjetivo? Se podría hacer alguna divulgación sobre si los datos eran privados o públicos, y si incorporaban dimensiones que, de ser divulgadas, tendrían implicancias de privacidad personal.

3) Modelo: en sí mismo, así como el proceso de modelamiento, también podría hacerse transparente hasta cierto punto. Es muy importante saber qué variables o características utiliza el modelo como entrada. A menudo, esas características se ponderan: ¿cuáles son esas ponderaciones? Si los datos de entrenamiento del modelo se usaron en algún proceso de aprendizaje automático, sería pertinente caracterizar los datos usados para eso a lo largo de todas las dimensiones potenciales descritas con anterioridad. Algunas herramientas de modelamiento de software tienen diferentes suposiciones o limitaciones: ¿Cuáles fueron las herramientas utilizadas para hacer el modelo? Por supuesto, todo esto se relaciona con la participación humana, por lo que se necesita conocer la justificación de las ponderaciones y el proceso de diseño para considerar modelos alternativos o comparaciones de modelos. ¿Cuáles son los supuestos (estadísticos o de otro tipo) detrás del modelo y dónde surgieron esos supuestos?

4) Inferencia: las inferencias hechas por un algoritmo, como clasificaciones o predicciones, a menudo dejan dudas sobre la precisión o la posibilidad de error. Los creadores de algoritmos debieran considerar la evaluación comparativa con conjuntos de datos estándar y con medidas estándar de precisión para revelar algunas estadísticas clave. ¿Cuál es el margen de error? ¿Cuál es la tasa de precisión y cuántos falsos positivos y falsos negativos hay? ¿Qué medidas se toman para remediar los errores detectados?

5) Presencia algorítmica: quienes utilizan los algoritmos deberían revelarlo, en particular si se está utilizando una personalización del algoritmo. Se debería generar un proceso de evaluación continua del algoritmo.

Tal vez sea necesario crear nuevos roles multidisciplinarios relacionados con el modelamiento de riesgo algorítmico o el modelamiento de transparencia para que estas preguntas reciban una atención dedicada y sostenida. Se podrían buscar regulaciones que obliguen a la divulgación de información o al menos auditorías de rutina en torno a decisiones clave influenciadas algorítmicamente, como por ejemplo la calificación crediticia de individuos.

Finalmente, la propuesta de protección relacionada con generar una transparencia algorítmica debe ir de la mano con un cambio en la regulación. El primer paso sería incluir el derecho de protección ante sesgos algorítmicos como un derecho fundamental en la Constitución. Complementariamente, se debería crear una institucionalidad que resguarde la protección contra sesgos algorítmicos. Esta institucionalidad podría estar incluida en el Proyecto de Ley que regula la protección y el tratamiento de datos personales y crea la Agencia de Protección de Datos Personales (Boletines N°11.144-07 y 11.092-07, refundidos). Esta Agencia tendría por objeto velar por la efectiva protección de los derechos que garantizan la vida privada de las personas y sus datos personales, de conformidad a lo establecido en la ley, y fiscalizar el cumplimiento de sus disposiciones, incluido en el número 376 del borrador del consolidado de normas aprobadas para la propuesta constitucional realizada por el pleno de la convención constitucional (actualizado a agosto del 2022). Lo anterior podría significar un gran paso para abordar esta problemática, que tratado de manera multidisciplinaria como se señaló con anterioridad y con el foco de prevenir la ocurrencia de estos sesgos, junto con la correspondiente reparación en caso de ocurrencia, podría ser un complemento robusto para la protección contra los sesgos algorítmicos.

En relación con la propuesta de protección sugerida en este trabajo, respecto al derecho a la protección contra los sesgos algorítmicos, podemos verificar 3 posibles escenarios en relación con la vía en que esta propuesta se podría concretar: 1) vía texto definitivo de la propuesta de nueva constitución, 2) vía modificación a Ley 21.383 que modifica la carta fundamental actual, para establecer el desarrollo científico y tecnológico al servicio de las personas, 3) vía modificación al Proyecto de Ley que regula la protección y el tratamiento de los datos personales y crea la Agencia de Protección de Datos Personales. (Número de boletín 11144-07 refundido con boletín 11092-07).

Las sugerencias de modificación a los respectivos cuerpos legales se detallan a continuación:

Modificación al texto definitivo propuesta nueva constitución, Artículo 376, crea la Agencia Nacional de Protección de Datos

Se sugiere incorporar la protección contra sesgos algorítmicos de acuerdo con la siguiente indicación:

“Existirá un órgano autónomo denominado Agencia Nacional de Protección de Datos, que velará por la promoción y protección de los datos personales, junto con la protección ante eventuales sesgos algorítmicos que deriven del tratamiento de datos, con facultades de normar, investigar, fiscalizar y sancionar a entidades públicas y privadas, el que contará con las atribuciones, la composición y las funciones que determine la ley.”

Modificación a Ley 21383 que modifica la carta fundamental, para establecer el desarrollo científico y tecnológico al servicio de las personas.

"Artículo único. - Modifícase el número 1º del artículo 19 de la Constitución Política de la República, de la siguiente forma:

Se sugiere incorporar la protección contra sesgos algorítmicos de acuerdo con la siguiente indicación:

1) Reemplázase, en el actual párrafo final, el punto y coma por un punto y aparte.

2) Agrégase el siguiente párrafo final, nuevo:

"El desarrollo científico y tecnológico estará al servicio de las personas y se llevará a cabo con respeto a la vida y a la integridad física y psíquica. La ley regulará los requisitos, condiciones y restricciones para su utilización en las personas, debiendo resguardar especialmente la actividad cerebral, así como la información proveniente de ella; también se protegerá ante eventuales sesgos algorítmicos que deriven del tratamiento de datos, de acuerdo a lo establecido en la ley para estos efectos".

Modificación al Proyecto de Ley que regula la protección y el tratamiento de los datos personales y crea la Agencia de Protección de Datos Personales. (Número de boletín 11144-07 refundido con boletín 11092-07).

Respecto al Título I de los derechos del titular de datos personales, se sugiere agregar nuevo artículo de acuerdo con lo siguiente:

“Derecho a la protección contra sesgos algorítmicos. El titular de datos tiene derecho a la protección contra sesgos algorítmicos derivado del tratamiento de los datos. La ley establecerá los requerimientos relativos al proceso de denuncia ante la ocurrencia de sesgos algorítmicos, junto con las posibles acciones reparativas de acuerdo a lo establecido en la ley. La Agencia deberá fiscalizar e investigar la ocurrencia de estos sesgos algorítmicos, de acuerdo a las atribuciones que la ley indique.”

En relación con las atribuciones que debería tener la Agencia de Protección de Datos Personales, ante la ocurrencia de sesgos algorítmicos, la Agencia debería operar en base a denuncias realizadas por los afectados. Una vez acogida la denuncia la Agencia debería llevar a cabo una investigación, en base a los antecedentes presentados por los afectados por los sesgos algorítmicos.

En cuanto al proceso investigativo respecto al análisis de la ocurrencia del sesgo algorítmico, este se sugiere que debería llevarse a cabo a través de un proceso de transparencia algorítmica, el cual fue detallado al comienzo de la propuesta de protección del presente trabajo. La finalidad principal es determinar las causas de la generación del sesgo algorítmico y de esta forma generar las posibles correcciones y reparaciones hacia los afectados.

De acuerdo con el “Párrafo Primero De la responsabilidad, las infracciones y las sanciones aplicables a las personas naturales o jurídicas de derecho privado”, incluido en el proyecto de ley, se establecen las infracciones leves, graves y gravísimas, que para el caso de los sesgos algorítmicos y dependiendo de la gravedad del daño generado por el sesgo, es que deberían establecerse sanciones, las cuales deberían estar incluidas en el artículo 35 del proyecto de ley, “Sanciones”.

Respecto a las partes, denunciante (quién se vio afectado por el sesgo algorítmico) y denunciado (quien es el responsable de la generación del sesgo algorítmico), ambas podrían participar de una primera instancia de resolución del conflicto, generando las reparaciones y compensaciones respectivas de acuerdo a lo que indique la ley y la capacidad resolutoria de la Agencia. En el caso de no existir un acuerdo reparatorio y/o compensatorio entre las partes, se podría llevar el caso a una segunda instancia, que podría estar determinada en la justicia civil, en donde la Agencia de Protección de Datos Personales actuaría como organismo denunciante junto con la parte afectada por el sesgo algorítmico. Es en esta segunda instancia donde el tribunal pertinente podría solicitar mayores antecedentes a las partes, pudiendo generar una sentencia definitiva en la cual ambas partes podrían apelar de acuerdo con las disposiciones establecidas por el tribunal respectivo.

VI. CONCLUSIONES

En la contingencia actual respecto a la automatización del tratamiento de la información, específicamente en relación con la utilización de algoritmos que reemplacen tareas realizadas manualmente por humanos, es innegable que la cantidad de datos que estos algoritmos pueden procesar ha crecido exponencialmente en los últimos años. Dado lo anterior, es donde existe la posibilidad que se generen sesgos algorítmicos en el tratamiento de los datos. Es especialmente relevante que se puedan suscitar posibles situaciones de discriminación, ya sea de manera intencionada o no, en donde los derechos de las personas que se vean afectados por estos sesgos

algorítmicos, en donde las víctimas puedan quedar en una situación desfavorable y sin ningún tipo de reparo.

Como se revisó, es necesario determinar distintas fuentes de protección en relación con el origen del sesgo y generar las respectivas reparaciones hacia los usuarios afectados. Los estudios recientes indican que la institucionalidad relacionada con esta materia actualmente carece de un tratamiento especial a las situaciones generadas por sesgos algorítmicos, por lo que con el avance de las tecnologías actual se hace necesario contar con una protección no sólo a los derechos humanos universales, sino que también se debe crear una protección a los derechos digitales, los cuales deberían tener un tratamiento más especializado respecto a las temáticas en que los derechos de las personas se vean vulnerados en el ámbito digital. Es necesario poner especial atención en el sesgo algorítmico que más se menciona en la literatura, el que está relacionado con cualquier tipo de discriminación que atente contra los derechos de las personas.

El caso del proyecto de ley chileno significa un gran avance en esta materia, el cual a nivel mundial tiene una posición aventajada respecto a los demás países en materia legislativa respecto a la protección de los neuroderechos, sin embargo, si bien el proyecto de ley chileno contempla varios aspectos importantes respecto al acceso a las tecnologías de información y tratamiento de los datos de las personas, deja vulnerable la protección ante sesgos algorítmicos que sistemas automatizados puedan generar. Es importante el tratamiento de estos sesgos a nivel gubernamental, en donde se pueda establecer una institucionalidad que a la par con la legislación puedan proteger de mayor y mejor manera situaciones generadas por sesgos algorítmicos. Se debe avanzar hacia una auditabilidad en la utilización de algoritmos de manera tal que exista transparencia en la composición de este, relacionado solo con las variables que puedan ser utilizadas para la construcción del algoritmo y sean susceptibles a generar sesgo.

Paralelamente, es de importante relevancia el proyecto de ley chileno que pretende crear la Agencia de Protección de Datos Personales, la cual tendrá por objeto velar por la efectiva protección de los derechos que garantizan la vida privada de las personas y sus datos personales, y fiscalizar el cumplimiento de sus disposiciones contenidas en el proyecto de ley. Lo anterior podría tener una estrecha relación con la protección de las personas ante los sesgos algorítmicos, en caso de que este neuroderecho no se incluya como una garantía constitucional a ser protegida. Lo importante es que, independiente del órgano del Estado desde el cual emane la protección, esta se materialice y se perfeccione a medida que aumente la disponibilidad y/o demanda de tecnologías que hacen uso de algoritmos, en particular en el ámbito de la toma de decisiones.

REFERENCIAS

- Adadi, A., Berrada, M. (2018). Peeking inside the black-box: A Survey on explainable artificial intelligence (XAI).
- Angwin, J., Larson, J., Mattu, S., Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. and it's biased against blacks, ProPublica, 2016.
- Araya, C. (2021). Transparencia algorítmica ¿un problema normativo o tecnológico?, revista CUHSO, vol. 31.
- Baeza-Yates, R., Peiró, K. (2019). ¿Es posible acabar con los sesgos de los algoritmos? Recuperado de [DCC Universidad de Chile](#)
- Baker, R., Hawn, A. (2021). Algorithmic bias in education. Journal of Artificial Intelligence in Education, 2021.
- Banu, A., Dau, N., Iren, D. (2020). Preventing algorithmic bias in the development of algorithmic decision-making systems: A delphi study. 53rd Hawaii International Conference on System Sciences (HICSS), 2020.
- Benítez, L. (2019). Ética y transparencia para la detección de sesgos algorítmicos de género. Estudios sobre el mensaje periodístico, 25(3) 2019: 1307-1320
- Biblioteca del Congreso Nacional de Chile. (2021). Ley 21.383. Recuperado de [BCN](#)
- Boletín N° 13.828-19. Proyecto de ley sobre protección de los neuroderechos y la integridad mental, y el desarrollo de la investigación y las neurotecnologías. (2020) Recuperado de [Diario Constitucional](#)
- Bozdag, E. (2013). Bias in algorithmic filtering and personalization. Springer Science+Business Media Dordrecht.
- Burk, D. (2019). Algorithmic fair use. The University of Chicago Law Review 2019, Vol. 86, No. 2, pp. 283-308
- Carta de Derechos Digitales. (2021). Recuperado de [Gobierno de España](#)
- Consolidado normas aprobadas para la propuesta constitucional por el pleno de la convención: texto definitivo (2022). Recuperado de [Convención Constitucional](#)
- Danks, D., London, A. (2017). Algorithmic bias in autonomous systems. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence.
- Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. Recuperado de [Reuters](#)
- Datta, A., Tschantz, M.C. (2015). Automated experiments on ad privacy. A tale of opacity, choice and discrimination. Proceedings on Privacy Enhancing Technologies, 2015(1), 92 – 112.

Diakopoulos, N., (2016). Accountability in algorithmic decision making. *Communications of the ACM*, vol. 59, N°2, pp 56-62.

Fuller, J., Raman, M., Sage-Gavin, E., Hines, K. (2021). Hidden workers: untapped talent. Harvard Business School.

Hajian, S., Bonchi, C., Castillo, C. (2016). Algorithmic bias: From discrimination discovery to fairness-aware data mining. *KDD 2016*.

Jackson, J. (2018) Algorithmic bias. *Journal of leadership, accountability and ethics*, vol 15(4) 2018, pp. 55-65.

Kleinberg, J., Ludwig, J., Rambachan, A., Mullainathan., S. (2018). Algorithmic fairness. *The American Economic Association*, vol 108.

Moya, G., Le, V. (2021). Algorithmic Bias explained: How automated decision-making becomes automated discrimination. The Greenlining Institute, 2021.

Müller, O., & Rotter, S. (2017) Neurotechnology: Current developments and ethical issues. *Frontiers in systems neuroscience*, 2017. Recuperado de [National Library of Medicine](#)

Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan., S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Journal of Science (New York, N.Y.)*, 2019.

Real Academia Española. Recuperado de [Definición de sesgo](#)

Suresh, H., & Guttag, J. V. (2020). A framework for understanding unintended consequences of machine learning. *Equity and access in algorithms, mechanisms, and optimization*. Recuperado de [Cornell University](#)

Yuste, R., Genser, J., Herrmann, S. (2021). It's time for neuro-rights. *Journal of International Relations and Sustainable Development*, 2021.