

ARTÍCULO DE INVESTIGACIÓN

## Transparencia algorítmica ¿un problema normativo o tecnológico?

*Algorithmic transparency - a regulatory or a technological problem?*

CARLOS ARAYA PAZ

*Investigator independiente, Chile*

**RESUMEN** Con la evolución de las tecnologías, cada vez va a ser más frecuente que ciertas decisiones tales como selección de personal, otorgamiento de crédito o selección de parejas, sean adoptadas por algoritmos de aprendizaje automático. Muchas de esas decisiones pueden contener sesgos que derivan en decisiones discriminatorias que afectan al sujeto sobre el que recae dicha decisión.

Para revelar cómo se adoptaron dichas decisiones, se ha recurrido a la idea de transparencia en materia de algoritmos, implementándola a través de regulaciones. Dicho enfoque normativo o regulatorio ignora una premisa básica: la transparencia no es solamente un problema legal, sino también un problema tecnológico, que adquiere relevancia cuando hablamos de algoritmos de aprendizaje automático.

Los algoritmos de aprendizaje automático constituyen la base para el desarrollo de una tecnología idónea para automatizar procesos más sofisticados que el enfoque clásico de programación, pues supone acceder a reglas más potentes y flexibles, que se van ajustando automáticamente al entorno a medida que los datos de la muestra van evolucionando y reconociendo variaciones en función de muchísimos factores. Sin embargo, el grado de sofisticación de estos algoritmos conlleva una complejidad que, a su vez, incide en la dificultad para su interpretación.

El enfoque normativo desconoce la complejidad antes mencionada, generando un marco regulatorio insuficiente y rígido. El presente artículo propone superar dicha visión a través de propuestas flexibles, comprendiendo el rol de la tecnología, reconociendo su valor y estimulando su uso. En definitiva, propuestas que fomenten y fortalezcan la innovación en Chile.

**PALABRAS CLAVE** Transparencia; algoritmo; sesgos; aprendizaje automático; regulación.

**ABSTRACT** As technology evolves, it will become increasingly frequent for decisions such as personnel selection, credit granting or partner selection, to be adopted by machine learning algorithms. Many of these decisions may contain biases that result in discriminatory decisions that affect the subject on whom the decision falls.

In order to reveal how such decisions have been made, the idea of transparency in terms of algorithms has been used, implementing it through regulations. Such normative or regulatory approach ignores a basic premise: transparency is not only a legal problem, but also a technological problem, which becomes relevant when we talk about machine learning algorithms.

Machine learning algorithms are the basis for the development of a more sophisticated technology in automating processes than the classic programming approach, since it involves access to more powerful and flexible rules, which are automatically adjusted to the environment as the sample data evolve and recognize variations depending on many factors. However, the degree of sophistication of these algorithms leads to a complexity of their models, which generates a difficulty in their interpretation.

The regulatory approach does not take into account the complexity mentioned above, what generates an insufficient and rigid regulatory framework. This article proposes to exceed this vision through flexible proposals, understanding the role of technology, recognizing its value and stimulating its use, this means proposals that promote and strengthen innovation in Chile.

**KEYWORDS** Transparency; algorithm; bias; machine learning; regulation.

## Introducción

En el 2014, Amazon lideró una iniciativa para la selección de personal basado en un sistema computacional. La iniciativa tuvo que suspenderse, dado que favorecía exclusivamente a postulantes hombres en desmedro de candidatas. La opinión pública responsabilizó al algoritmo, sin embargo, éste había sido preparado con la base de datos del personal de Amazon que, en su mayoría, eran ingenieros hombres (Sandri, 2019). En el 2016, la iniciativa Sesame Credit, creada por Alibaba para el mercado chino, generó polémica al efectuar calificación crediticia basada en métricas sociales, en la que los ciudadanos chinos no solo se veían afectados por su comportamiento crediticio, el uso de tarjetas de crédito y el pago regular de sus deudas, sino que además -en virtud del algoritmo de la plataforma- en base a su comportamiento en redes

sociales o el tipo de compras que hacen online (Kobie, 2019). En el 2018, se presentó una demanda contra Facebook por discriminar a mujeres, veteranos de guerra discapacitados y madres solteras, por la forma en que permitía a los anunciantes dirigirse a los consumidores a través de anuncios (Bagli, 2018). En noviembre de 2019, David Heinemeier Hansson, comentó cómo su mujer, había solicitado la misma Apple Card que él, obteniendo un límite de crédito muy inferior al suyo, a pesar de que declaraban impuestos conjuntamente y que, además, ella tenía mejor historial crediticio que él (Heinemeier Hansson, 2019). En el 2019, la doctora Melisa Hamilton publicó un estudio en el que determinó que el software de gestión de casos y apoyo a la toma de decisiones utilizada por los tribunales de justicia norteamericanos, COMPAS, tenía sesgos negativos en relación con imputados de origen latino (Butler, 2019).

Estos son algunos casos que ilustran la importancia creciente que están adquiriendo los modelos de *machine learning* (en adelante “aprendizaje automático”), y particularmente, las decisiones que ellos adoptan sobre la base de los algoritmos con los que estos sistemas se construyen. Es la prueba irrefutable que nos dirigimos hacia una sociedad en la que cada vez más elementos de nuestras vidas van a ser evaluados por un algoritmo. El problema radica -tal como se señaló en los ejemplos precedentes- en que los datos que alimentan a dichos algoritmos pueden contener sesgos que inciden en la decisión tomada, lo que finalmente impacta en los individuos y en sus derechos.

La eventual afectación de derechos es lo que un tribunal de los Países Bajos tuvo en consideración, en fallo de fecha 5 de febrero de 2020, al declarar contrario al Convenio Europeo de Derechos Humanos y, por tanto, ilegal, el uso de algoritmos diseñados para combatir el fraude social (Henley & Booth, 2020).

Una fórmula a la que se ha recurrido para resolver el problema de los sesgos, desde la perspectiva de las políticas públicas, es abogar por una transparencia en la toma de decisiones basadas en el uso de algoritmos. Transparencia que, en el contexto que nos convoca, toma el nombre de “*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 (Sangüesa, 2018).

La necesidad de transparencia en la toma de decisiones ha sido de tal entidad, que los gobiernos la han puesto como prioridad en sus respectivas agendas. Esta necesidad de transparencia ha sido recogida por la Comisión Europea, la Organización para la Cooperación y el Desarrollo Económico (OCDE), y recientemente por Estados Unidos.

En Chile, el tema ha cobrado importancia y a la fecha, existen dos iniciativas en la materia. La primera fue el documento elaborado por la Comisión Desafíos del Futuro, Ciencia, Tecnología e Innovación del Senado, que precisamente situó la transparencia y seguridad como uno de sus ejes (Senado de Chile, 2019). La segunda se

refiere al encargo que el Presidente de la República, realizó al Ministerio de Ciencia, Tecnología, Conocimiento e Innovación para el desarrollo de una Política Nacional de Inteligencia Artificial, en que se contengan los lineamientos estratégicos que deba seguir el país en esta materia, entre los que destaca la transparencia en materia de algoritmos (Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, 2020).

Lo anterior pone de relieve la importancia de la transparencia, especialmente en la toma de decisiones basadas en algoritmos. Sin embargo, su implementación, dista de ser sencilla dada la creciente complejidad de la tecnología sobre la que recae la pretendida transparencia. En ese sentido, el enfoque normativo actual ha sido *top-down*, es decir, como una imposición regulatoria que debe ser obedecida por las empresas que trabajan con algoritmos.

El presente artículo busca analizar dicho problema y proponer políticas públicas equilibradas y flexibles, concebidas desde el seno de la misma tecnología. Sobre este punto cabe precisar que el objetivo de este artículo es proponer políticas públicas desde la vereda regulatoria, no un artículo científico sobre algoritmos. Sin perjuicio de ello, se abordarán aspectos técnicos para una mejor comprensión de la materia. Para dichos fines, este artículo se estructura de la siguiente forma. En la primera sección, estableceremos el marco conceptual sobre el que se estructura este trabajo, abordando de forma general, la transparencia en la toma de decisiones humanas y su complejidad. Posteriormente, efectuaremos una breve digresión para entender conceptos como inteligencia artificial y aprendizaje automático. Una vez sentadas las bases de dichos conceptos, analizaremos la transparencia en la toma de decisiones efectuadas por algoritmos y los problemas que de ella derivan. En la segunda sección, fijado el marco conceptual del tema que nos convoca, examinaremos soluciones propuestas desde la perspectiva normativa y sus principales críticas. En la tercera sección, habiendo estudiado lo anterior, propondremos el diseño de políticas públicas que puedan resolver de mejor forma los desafíos que plantea la transparencia algorítmica. Finalmente, en la cuarta sección, concluiremos los aspectos relevantes sobre la transparencia algorítmica.

## **Nociones generales sobre la transparencia**

### *Toma de decisiones humanas y su transparencia*

El supuesto implícito que subyace al llamado a transparencia es la idea que entendiendo cómo funciona internamente un sistema, las personas no solo se sentirán más seguras y confiadas en el funcionamiento del mismo, sino que, además, les otorgará un control sobre éste, permitiéndoles asignar responsabilidades en caso de que su funcionamiento no sea correcto.

La transparencia, por lo tanto, refuerza la percepción de confianza y seguridad que tienen los individuos en el funcionamiento de sistemas complejos.

Sin perjuicio de lo anterior, bajo ciertas circunstancias, implementar la transparencia puede resultar un desafío dada la complejidad que conllevan algunos sistemas. Uno de los sistemas más complejos -por excelencia- es el cerebro humano.

El cerebro humano es un órgano infinitamente difícil de comprender, una verdadera maquinaria dotada de millones de neuronas que, a su vez, detentan miles de millones de conexiones. Es dicha complejidad la que debemos considerar en dos fases para los efectos de este artículo: en la toma de decisiones y en las explicaciones a las decisiones tomadas.

La toma de decisiones en los humanos es un proceso altamente complejo. Dicha decisión dependerá en gran medida de la información que se tiene sobre los hechos a decidir: si tenemos un mayor grado de información, la decisión se puede tomar con mayor certeza que en el evento en que desconocemos la información previa. En estos casos, entran una serie de factores, no siempre racionales, sino que a veces, elementos sociales, la experiencia propia, las emociones, la intuición, etc<sup>1</sup>. En otras palabras, no siempre opera un razonamiento lógico y consecuencial en esta fase. Decisiones como volver con una expareja o patear un penal en un partido de fútbol, pueden revestir un grado de complejidad inaudito.

Tomada una decisión, puede que se solicite una explicación sobre por qué fue adoptada dicha decisión. En esta fase, ya no opera la intuición o las emociones, sino la racionalidad, a través de historias que llegan a nuestra mente para explicar los motivos que subyacen en una decisión. El objetivo de dicha explicación es contar una historia que sea aceptable para quien solicita la explicación, y para que sea robustamente verosímil, dicha historia tiene que tener algunos elementos de verdad. Sin embargo, no siempre logramos explicar con eficacia la decisión tomada. Supongamos una decisión tomada sobre la base de la intuición, ¿cómo explicamos una intuición? En el fondo, hay decisiones que son difíciles de explicar racionalmente para un ser humano, y no por ello, resultan inválidas.

---

1. En esta etapa pueden operar distintos sesgos. Los más conocidos son: (a) El sesgo estadístico (aquel que procede de cómo obtenemos los datos, de errores de medida o similares, por ejemplo, si la policía está presente en algunos barrios más que en otros, no será extraño que la tasa de criminalidad sea más alta donde tenga mayor presencia); (b) 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. Los estereotipos de las personas de un país son un ejemplo claro); y (c) El sesgo cognitivo (aquel que nos identifica y que depende de nuestra personalidad, de nuestros gustos y miedos, por ejemplo, si leemos una noticia que está alineada con lo que pensamos, nuestra tendencia será validarla, aunque sea falsa. Esta última desviación se conoce como 'sesgo de confirmación'. Buena parte de las noticias falsas (fake news) se alimentan de este razonamiento para difundirse más rápidamente. Por este motivo, si no nos cuestionamos lo que leemos o vemos, corremos el riesgo de avanzar hacia una involución humana) (Baeza-Yates & Peiró, 2019).

Es el cerebro humano el concepto que se tiene como referencia al hablar de inteligencia artificial, y más particularmente, aprendizaje automático, y que analizaremos en la siguiente sección.

### *Inteligencia artificial y aprendizaje automático*

Inteligencia Artificial (en adelante “IA”) es la ciencia de construir máquinas que realicen tareas complejas que serían catalogadas como “inteligentes” si fueran realizadas por humanos. Esto trae como consecuencia que un sistema inteligente puede ser concebido bajo dos paradigmas, uno asociado a las capacidades de razonamiento y abstracción, y otro relacionado a las capacidades de actuación (Norvig & Rusell, 2004, p. 2). Esto origina cuatro paradigmas de IA: (1) Sistemas (de IA) que *piensan* como humano, (2) Sistemas que *actúan* como humanos, (3) Sistemas que *piensan* racionalmente y (4) Sistemas que *actúan* racionalmente cuyos detalles se muestran en la Tabla 1. (Norvig & Rusell, 2004, p. 2). La utilización de un paradigma u otro depende del objeto de estudio y de la capacidad que deseamos dotar a un sistema.

Tabla 1. Algunas definiciones de inteligencia artificial, organizadas en cuatro categorías.

<b>Sistemas que piensan como humanos</b>	<b>Sistemas que piensan racionalmente</b>
<p>“El nuevo y excitante esfuerzo de hacer que los computadores piensen...máquinas con mentes, en el más amplio sentido literal”. (Haugeland, 1985).</p> <p>“[La automatización de] actividades que vinculamos con procesos de pensamiento humano, actividades como la toma de decisiones, resolución de problemas, aprendizaje...” (Bellman, 1985).</p>	<p>“El estudio de las facultades mentales mediante el uso de modelos computacionales”. (Charniak y McDermonnt, 1985).</p> <p>“El estudio de los cálculos que hacen posible percibir, razonar y actuar”. (Winston, 1992).</p>
<b>Sistemas que actúan como humanos</b>	<b>Sistemas que actúan racionalmente</b>
<p>“El arte de desarrollar máquinas con capacidad de realizar funciones que cuando son realizadas por personas requieren de inteligencia”. (Kurzweil, 1990).</p> <p>“El estudio de cómo lograr que los computadores realicen tareas que, por el momento, los humanos hacen mejor”. (Rich y Knight, 1991).</p>	<p>“La Inteligencia Computacional es el estudio del diseño de agentes inteligentes” (Poole et al., 1998)</p> <p>“IA...está relacionada con conductas inteligentes en artefactos”. (Nilsson, 1998).</p>

Debido a estos paradigmas, la IA es una ciencia multidisciplinaria que involucra teorías y modelos de las Ciencias Cognitivas, Lingüística, Neurociencias, Matemáticas, Informática y Biología, entre otras. Por otro lado, con el fin de desarrollar sistemas inteligentes que utilicen capacidades cognitivas cercanas a las humanas, la IA se basa en el trabajo de varias ramas tales como la Representación de Conocimiento, Aprendizaje Automático (“Machine Learning”), Visión Artificial, Robótica, Agentes Autónomos, Planificación y Razonamiento Automático, Búsqueda y Resolución de Problemas, y Procesamiento de Lenguaje Natural. Un aspecto común de todas estas ramas científicas es que involucran procesos de toma de decisiones automáticos y autónomos (sin intervención humana), lo cual conlleva dos problemas mayores:

- *Transparencia de los modelos computacionales generados*: a pesar de que el número de aplicaciones de IA exitosas en forma transversal en muchos dominios ha crecido de forma exponencial, las decisiones en algunos casos son bastante opacas y por tanto comprensibles solamente a los científicos o diseñadores de los sistemas.
- *Sesgos de los métodos y/o conocimiento utilizado por los modelos previos*: en algunas situaciones, los sistemas de IA son “alimentados” con experiencia o conocimiento humano, para posteriormente tomar decisiones autónomamente. Sin embargo, esto puede traer consigo sesgos importantes de género, raza, etario, etc. que pueden ser perjudiciales a las personas.

Una de las áreas de la IA se conoce como *Planificación Automática*. Ésta persigue razonar automáticamente para determinar la “secuencia” de acciones posibles que debe tomar un sistema para ir desde una situación final a una situación deseada (i.e., imagine que un robot debe decidir automáticamente, las acciones a realizar para ir de un punto A hacia un punto B), y en el cual la “salida” de este razonador es el “plan” que debe seguir el sistema. En este caso, el plan es explícito y puede ser bastante comprensible, por lo que no existen mayores problemas de transparencia en las decisiones. Sin embargo, en otras áreas de la IA tales como el aprendizaje automático, existen variadas tecnologías cuyos modelos para la toma de decisiones no son transparentes, ni se comprende la forma en que se tomó una decisión. En términos simples, las técnicas de aprendizaje automático permiten que la máquina (computador) aprenda a realizar tareas automáticamente mejorando con la experiencia (Shalev-Shwartz & Ben-David, 2014, p. 19). Esto es fundamental pues el enfoque clásico de “programación” de los computadores asume que la máquina tiene un conjunto de instrucciones (i.e., programa) que especifica paso a paso lo que debe realizar para cumplir un cierto objetivo, mientras que el aprendizaje automático se caracteriza por abordar problemas complejos en los cuales no es posible tener secuencias de pasos bien delimitadas ni claras, por lo que se debe dotar al computador de la facultad de “aprender” automáticamente a realizar alguna tarea a partir de ejemplos o la experiencia, o sea, sin ser programados explícitamente (Galeano, 2019, p. 20).

En ese orden de ideas, los sistemas de IA que utilizan técnicas de aprendizaje automático trabajan sobre la base de métodos computacionales que le permiten aprender “modelos”, que posteriormente se pueden utilizar para realizar tareas altamente complejas. Estos métodos computacionales son algoritmos de difícil comprensión que no le indican a la máquina los pasos del problema (que se desconocen) sino la forma en que puede aprender a partir de ejemplos o experiencia pasada. Tradicionalmente, un algoritmo es un “*conjunto ordenado y finito de operaciones que permite hallar la solución de un problema*” (Real Academia de la Lengua Española, 2020). Lo anterior implica una descripción precisa de pasos a seguir para alcanzar la solución a un problema dado. Usando una analogía, un algoritmo es como una receta, una serie de instrucciones que nos dicen cómo trabajar mecánicamente con ciertos conjuntos de datos. Para lograr dicha solución, los algoritmos reciben una entrada (*input*), posteriormente procesan esa entrada (sistema de aprendizaje), y tras un tiempo finito, entregan como respuesta un resultado determinado (*output*).

Dependiendo de la naturaleza del problema a resolver, el dominio, y el tipo de conocimiento disponible, las técnicas de aprendizaje automático pueden ser de diferentes clases, generando diversos tipos de modelos (i.e, imagine un modelo como una función que posee parámetros que se desconocen y que se deben “aprender”). En general, estos se pueden dividir en tres grandes métodos:

(1) *Métodos de aprendizaje automático supervisado* (i.e, requieren ejemplos o supervisión para “generar” los modelos). Esto incluye técnicas como árboles de decisión, redes neuronales artificiales, métodos de inferencia probabilística, métodos de aprendizaje profundo, etc.

(2) *Métodos de aprendizaje automático no-supervisado* (i.e., no se posee ejemplos por lo que deseamos aprender relaciones o patrones que pueden estar expresando mis datos o conocimiento). Esto incluye técnicas de “agrupamiento” o clustering, extracción de patrones de asociación, etc.

(3) *Métodos de aprendizaje automático semi-supervisados* (i.e., no se poseen muchos ejemplos para “entrenar” a la máquina, por lo que gradualmente se debe ir preparando y corrigiendo el modelo a partir de pocos casos iniciales de “muestra”).

En el caso particular de los métodos de aprendizaje automático supervisado, en general las tareas tienen que ver con problemas de “predicción” (i.e., proyección de valores futuros de ciertas variables) de datos continuos (también llamados “Regresión”) o de datos discretos (también llamados “Categorización”). Estos modelos de predicción pueden, por ejemplo, mirar millones de informes de crédito y clasificar a las personas en categorías separadas de riesgo crediticio o procesar imágenes y separar aquellas que contienen rostros de las que no contienen rostros (Bathae, 2018, p. 900).

Los métodos computacionales que subyacen a las diferentes técnicas de aprendizaje automático poseen varios tipos de complejidad, incluyendo:

- 1) *Complejidad temporal*: estimación matemática de la eficiencia en tiempo de ejecución de un algoritmo.
- 2) *Complejidad espacial*: estimación matemática de la eficiencia en el uso del espacio en memoria de un algoritmo.
- 3) *Complejidad de diseño*: dificultad en la comprensión de la lógica que subyace al algoritmo de aprendizaje.

Este último aspecto de la complejidad puede incidir en que los métodos de aprendizaje automático sean verdaderas cajas negras para los humanos, esto es, que exista una incapacidad humana para comprender plenamente el proceso de toma de decisiones del sistema y para predecir las decisiones o resultados adoptados por éste (Bathae, 2018, p. 892). En otras palabras, a medida que los algoritmos de aprendizaje automático se hacen más y más sofisticados, la posibilidad de que una mente humana entienda los procedimientos que utiliza para llegar a un resultado determinado disminuyen progresivamente, lo que nos lleva, en último término, a tener una caja negra que genera resultados, sin entender cómo se llega a ellos. No obstante lo anterior, dicha consecuencia no es necesariamente nociva, un sistema de aprendizaje automático de caja negra puede hacer cosas de maneras que sus creadores no pudieron entender o ser capaces de predecir (Bathae, 2018, p. 907)<sup>2</sup>.

La opacidad antes mencionada, puede tener distintas causas como señalamos con anterioridad al examinar los tipos de complejidad de los modelos computacionales que subyacen a las técnicas de aprendizaje automático. Para enfatizar lo anterior, conviene tener presente dos ejemplos que ilustran sobre el origen de la mencionada opacidad: (i) la complejidad de la estructura del algoritmo, como en el caso de una red neuronal profunda, que consiste en miles de neuronas artificiales que trabajan juntas de manera difusa para resolver un problema; o (ii) el algoritmo de aprendizaje automático basado en relaciones geométricas que los humanos no pueden visualizar, como en el caso de las máquinas de vectores de soporte (SVM). Por consiguiente, en ambos casos la falta de transparencia no proviene de un capricho, sino que, de aspectos técnicos asociados a la complejidad y dimensionalidad, respectivamente, que hacen imposible que una persona pueda comprender dichos procesos (Bathae, 2018, p. 897).

---

2. Los sistemas de IA generalmente funcionan bajo la lógica "*outside-of-the-box*", referido al pensamiento no convencional o lógico, es decir, para la consecución de un resultado dado se sigue una secuencia de razonamiento que no resultaría lógica si dicha secuencia fuese ejecutada por un humano.

De esa forma, los modelos resultantes de los algoritmos de aprendizaje automático pueden ser muy difíciles de comprender en detalle, ya que pueden captar relaciones o patrones complejos y opacos entre las variables de lo que parecía ser un simple conjunto de datos (como los registros del resumen del historial financiero, crediticio y de empleo, de solicitantes de préstamos, junto con el resultado de sus préstamos). Así, un algoritmo simple aplicado a conjuntos de datos simples puede, no obstante, dar lugar a modelos inescrutables. De igual manera, algoritmos más complejos aplicados a conjuntos de datos aparentemente no relacionados pueden dar lugar a una opacidad del modelo aún mayor (Kearns & Roth, 2019).

La opacidad de estos algoritmos muchas veces tiene correlación con el perfeccionamiento de la representación, ajuste y rendimiento del modelo. Sin embargo, ello ocurre a costa de un grado de interpretabilidad sustantivamente menor y con un mayor tiempo dedicado al entrenamiento de dicho modelo. En efecto, cuando el modelo alcanza ese nivel de complejidad, la capacidad de correlacionar el resultado (*output*) con las variables introducidas en la entrada (*input*) se reducen significativamente. Por ello, lograr interpretar la causalidad resulta muchas veces incomprensible. Por consiguiente, demostrar que un método de aprendizaje automático no ha tomado una decisión, basado en un criterio eventualmente discriminatorio puede resultar extremadamente complejo desde la perspectiva técnica.

### *Exigencia de transparencia a algoritmos de aprendizaje automático*

Aunque es razonable demandar una mayor transparencia algorítmica, cabe considerar que estos sistemas de IA (especialmente aquellos que utilizan técnicas de aprendizaje automático) toman como referente -aunque aún lejano- al cerebro humano, y si las personas no pueden explicar con claridad por qué tomaron una decisión dada, ¿cómo se le puede exigir a un sistema que aún no logra imitar al cerebro humano, que explique con claridad una decisión tomada que ni siquiera un humano puede explicar?

Un resultado que sea discriminatorio es ciertamente indeseable, pero culpar al sistema que procesa los datos que sirven como *input* y tildarlo como una tecnología discriminatoria, es no entender los fundamentos básicos de lo que se critica.

Tal como señalamos en el apartado anterior, los sistemas de aprendizaje automático son alta y necesariamente complejos. Esto es una consecuencia lógica de la sofisticación de la tecnología con el transcurso del tiempo.

En el caso de los algoritmos de aprendizaje automático, existe un constante *trade-off* entre la capacidad de representación de un modelo y su interpretabilidad o explicabilidad. En otras palabras, mientras más complejo en su interpretación sea un modelo, mayor precisión existirá en sus resultados. De hecho, a diferencia de los modelos de simulación, los datos y/o conocimiento que se proporciona a los métodos

de aprendizaje automático pueden carecer de cualquier correlación lógica para que un humano los interprete. Un ejemplo de lo anterior es *LinkedIn*, que posee más de 100.000 variables asociadas a cada usuario, con las cuales se entrenan los modelos de ML. Muchas de ellas son variables directas extraídas desde los perfiles (i.e., edad) y otras son indirectas (i.e., tiempo que un usuario tarda en hacer clic, tiempo que un usuario pasa leyendo, texto que el usuario produce en un cuadro de texto). Estas variables bien pueden contener buenos “indicadores” o características predictivas (*features*) sobre comportamientos individuales. Sin embargo, el método que las utiliza carece de formas convincentes de mostrar claramente estas explicaciones para una interpretación humana significativa (Edwards & Veale, 2017, pp. 59-60).

Frente a los problemas que derivan de la opacidad de ciertos algoritmos de aprendizaje automático, así como los eventuales sesgos que éstos pueden conllevar, se han tratado de proporcionar soluciones desde distintas veredas (Kearns & Roth, 2019).

Sin perjuicio de lo anterior, en este artículo nos haremos cargo de las soluciones que se han propuesto desde el enfoque normativo, sus críticas y cómo se puede proponer una política pública saludable en esta materia.

### **Soluciones para el problema de la transparencia algorítmica: propuestas normativas existentes**

En esta sección analizaremos las propuestas normativas que han surgido en el contexto de la transparencia algorítmica, esto es, aquellas soluciones basadas en normas emanadas de una autoridad central y cuya inobservancia acarrea una sanción. A dichos efectos, analizaremos lo que ocurre en Estados Unidos y la Unión Europea, por ser éstos, bloques económicos fuertes y que han regulado o procurado regular la materia.

En Estados Unidos se presentó el año pasado un proyecto de ley denominado “*Algorithmic Accountability Act*”. El objetivo del proyecto de ley era obligar a las grandes empresas a analizar de forma periódica sus algoritmos y reparar aquellos que generen decisiones discriminatorias, injustas, sesgadas o imprecisas, lo anterior a través de un reglamento que emitiría la Comisión Federal de Comercio de Estados Unidos. Sin embargo, su ámbito de aplicación se restringe a las empresas que generan más de 50 millones de dólares al año, que poseen información sobre al menos 1 millón de personas o dispositivos, o que actúan principalmente como intermediarios de datos que compran y venden datos de consumidores. En caso de inobservancia, el Estado puede ejercer acciones civiles contra la empresa infractora (Booker, Wyden & Clarke, 2019). El proyecto de ley no ha estado exento de críticas, entre las que destacamos dos: (a) se refiere solo a los sistemas “*automatizados*” de toma de decisiones que -además- representan un alto riesgo, pero excluye la toma de decisiones igualmente riesgosas en las que ha intervenido o que provienen de un humano. De esa forma, el proyecto de

ley sugiere de forma implícita, que las decisiones automatizadas son intrínsecamente menos fiables o más peligrosas que las humanas, lo cual desincentiva la adopción de tecnologías como el aprendizaje automático, y sus consabidos beneficios sociales y económicos; y (b) se aplica solo a empresas con una alta facturación, sugiriendo de forma implícita que no es relevante -desde la óptica regulatoria- que una empresa de menor tamaño realice igualmente conductas discriminatorias que pueden tener un alto impacto en los consumidores.

Sin perjuicio de lo anterior, en el ámbito crediticio, Estados Unidos sí cuenta con la *“Equal Credit Opportunity Act”*, una normativa que sanciona al acreedor que discrimine a un solicitante -en una transacción crediticia- basándose para ello en aspectos como la raza, religión, nacionalidad, sexo, estado civil, y si tiene o no hijos, entre otros factores (Kagan, 2018). El problema de esta normativa radica en que: (a) solo se hace aplicable al ámbito crediticio, excluyendo la discriminación que pueda existir en otros campos; y (b) es una ley que data de 1974, y en definitiva, no provee herramientas para explicar en caso que el algoritmo funcione bajo una lógica de caja negra.

Examinado el estado de arte en Estados Unidos, conviene ahora enfocar nuestro estudio en otro de los grandes bloques económicos: la Unión Europea (en adelante “UE”). A diferencia de Estados Unidos, con su acotada regulación, la UE tiene un historial regulatorio sobre el tratamiento de datos personales que data de 1995, y que se ha actualizado y consolidado en lo que hoy se conoce como Reglamento General de Protección de Datos (“GDPR” por sus siglas en inglés). El GDPR es el reglamento europeo que regula lo relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos.

Sobre este punto, cabe destacar la importancia de los datos como variables de entrada (*input*) en la construcción de ciertos modelos de aprendizaje automático. Los datos que nos importan, a efectos del presente artículo, son los datos de carácter personal<sup>3</sup>, cualquier regulación que obstaculice el flujo o el tratamiento de dichos datos, impactará necesariamente en el *“output”*. De allí la importancia de la normativa que abarca dicho tema.

El GDPR resulta relevante para nuestro análisis por dos motivos: primero, por lo que se conoce como *“Efecto Bruselas”*, y segundo, por la regulación de las *“decisiones individuales automatizadas, incluida la elaboración de perfiles”* efectuada sobre la base de datos personales.

Un primer aspecto que considerar es el llamado *“Efecto Bruselas”*, éste se genera cuando la UE hace exigibles sus leyes y reglamentos a otras jurisdicciones, imponiendo requisitos normativos aplicables no solo a los miembros de la UE sino a los demás

---

3. Para los efectos de este artículo, definiremos como *“datos personales”* aquellos relativos a cualquier información concerniente a personas naturales, identificadas e identificables.

países que deseen concretar o expandir sus negocios en dicho bloque económico. Así, los países no miembros de la UE deben decidir: si modifican su legislación interna para adaptarse a la UE o simplemente cortan relaciones económicas con uno de los bloques financieros más atractivos a nivel mundial (Casey, Farhangi & Vogl, 2019, pp. 184-185). De forma tal que, frente a ese escenario, muchos países deciden modificar su normativa interna<sup>4</sup>. En efecto, el GDPR ha probado su fuerza expansiva en países como Israel, Nueva Zelanda, Argentina y Japón (Scott & Cerulus, 2018). De igual manera, países como Brasil, Australia, Corea del Sur y Tailandia han adaptado su normativa al GDPR (Simmons, 2019). Chile no ha sido la excepción. En efecto, si bien en la actualidad contamos con la Ley N°19.628 Sobre Protección a la Vida Privada que data de 1999, hoy se encuentra en discusión el Proyecto de ley (en adelante el “Proyecto”) que regula la protección y el tratamiento de los datos personales y crea la Agencia de Protección de Datos Personales (Boletín N° 11.144-07), tramitado en forma conjunta con el proyecto de ley sobre protección de datos personales (Boletín N° 11.092-07), y que busca adecuar la realidad chilena a estándares internacionales (GDPR) en materia de protección de datos personales.

Un segundo aspecto dice relación con la regulación de las decisiones individuales automatizadas, incluidas la elaboración de perfiles, efectuada sobre la base de datos personales. En este punto cobra relevancia, lo que la doctrina ha llamado el “*derecho a la explicación*”, que se sustenta en la interpretación de ciertas disposiciones del GDPR, en virtud del cual, las entidades que manejan datos personales de los ciudadanos de la UE deben garantizar “*un tratamiento leal y transparente*” de estos datos, para lo cual es necesario que los ciudadanos tengan acceso a “*información significativa sobre la lógica aplicada*” en determinados sistemas automatizados de adopción de decisiones (Casey et al., 2019, p. 154). En otras palabras, el derecho a la explicación -como su nombre lo indica- consiste en el derecho de recibir una explicación respecto del *output* de un determinado algoritmo, es decir, por qué un algoritmo efectuó una decisión que afectó a una persona. ¿Quién debe efectuar la explicación? El responsable del tratamiento de datos personales, es decir, toda persona natural o jurídica, pública o privada, que decide acerca de los fines y medios del tratamiento de datos personales. En otras palabras, la empresa u organismo que efectuó el tratamiento.

El núcleo del derecho a la explicación se encuentra en el artículo 22 del GDPR que establece lo siguiente:

---

4. Un ejemplo sobre este efecto es el GDPR, y el mecanismo normativo que utiliza para promover esta globalización jurídica es lo que se denomina “*nivel de protección adecuado*”, en virtud del cual, los países ajenos a la UE que deseen transferir datos personales desde o hacia la UE, deben contar con un nivel de protección normativa similar a los que gozan los miembros de la UE.

*“1. Todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar.*

*2. El apartado 1 no se aplicará si la decisión:*

*a) es necesaria para la celebración o la ejecución de un contrato entre el interesado y un responsable del tratamiento;*

*b) está autorizada por el Derecho de la Unión o de los Estados miembros que se aplique al responsable del tratamiento y que establezca asimismo medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado, o*

*c) se basa en el consentimiento explícito del interesado.*

*3. En los casos a que se refiere el apartado 2, letras a) y c), el responsable del tratamiento adoptará las medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado, como mínimo el derecho a obtener intervención humana por parte del responsable, a expresar su punto de vista y a impugnar la decisión.*

*4. Las decisiones a que se refiere el apartado 2 no se basarán en las categorías especiales de datos personales contempladas en el artículo 9, apartado 1, salvo que se aplique el artículo 9, apartado 2, letra a) o g), y se hayan tomado medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado.”*

Dicho artículo tiene su correlato en el artículo 8 bis que contiene el Proyecto en Chile, que establece lo siguiente:

***Artículo 8° bis.- Derecho de oposición a valoraciones personales automatizadas.***

*El titular de datos tiene derecho a oponerse a que el responsable adopte decisiones que le conciernan, basadas únicamente en el hecho de realizarse a través de un tratamiento automatizado de sus datos personales, incluida la elaboración de perfiles.*

*El titular no podrá ejercer este derecho de oposición en los siguientes casos:*

*a) Cuando la decisión del responsable sea necesaria para la celebración o ejecución de un contrato entre el titular y el responsable;*

*b) Cuando exista consentimiento previo y expreso del titular, y*

*c) Cuando lo disponga la ley.*

*En los casos de las letras a) y b) del inicio anterior, el responsable deberá adoptar las medidas necesarias para asegurar los derechos del titular, en particular el derecho a obtener intervención humana por parte del responsable, a expresar su punto de vista y a solicitar la revisión de la decisión.*

Ambos artículos comparten la misma base, sin embargo, en el caso chileno expresamente se excluye un “*derecho a la explicación*” y se lo reemplaza por un “*derecho de oposición*”. Esto quiere decir que, de aprobarse el Proyecto, en Chile no existiría una obligación legal de explicar cómo toma una decisión un algoritmo, sino solo un derecho a oponerse en el caso que una decisión sea basada únicamente en el tratamiento automatizado de sus datos personales<sup>5</sup>. Las críticas que derivan de ambas disposiciones resultan evidentes.

En primer término, ambos artículos se refieren a decisiones basadas “únicamente” en el tratamiento automatizado. De acuerdo con lo señalado en las Directrices del Grupo de Trabajo Sobre Protección de Datos del Artículo 29 (en adelante el “Directrices del Grupo del Artículo 29”)<sup>6</sup>, la decisión tiene que ser basada única y exclusivamente en el tratamiento automatizado de datos sin intervención o participación alguna de humanos en dicho proceso. Si existe participación humana y ésta es significativa -es decir, se lleva a cabo por una persona autorizada y competente para modificar la decisión- no se aplica el mencionado artículo 22 (Grupo de Trabajo Sobre Protección de Datos del Artículo 29, 2018, p. 22). El principal inconveniente que presenta dicha interpretación es que los sistemas de aprendizaje automático, que mayor impacto causan en la sociedad no son totalmente automatizados, sino que generalmente prestan un rol de apoyo a las decisiones que finalmente son adoptadas por humanos (Edwards & Vale, 2017, p. 45).

En segundo término, ambos artículos se refieren a “decisiones”. Las Directrices del Grupo del Artículo 29 no hacen referencia a este elemento, no obstante ello, resulta pertinente formular dudas que su interpretación podría generar. ¿Es realmente una decisión la que produce un sistema de aprendizaje automático o más bien es un *output* que sirve de base para la toma de una decisión? Cuando a un sistema de aprendizaje automático se le efectúa una consulta, éstos generan un *output* que constituye

---

5. La norma chilena mejora la redacción de la norma europea, porque se ha señalado, que el artículo 22 del GDPR establece una prohibición general de toma de decisiones individuales basadas en el tratamiento automatizado, pudiendo solamente procederse a dichas decisiones, bajo supuestos excepcionales, e incluso en esos casos, el responsable debe aplicar medidas para garantizar los derechos y libertades del interesado (Grupo de Trabajo Sobre Protección de Datos del Artículo 29, 2018, p. 21). En Chile, por el contrario, no se establece una prohibición general, sino un derecho a oponerse a dicho tratamiento. Es decir, en Chile se permite el tratamiento automatizado de datos personales, lo cual supone un avance en relación con la normativa europea.

6. Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679 del Grupo de Trabajo Sobre Protección de Datos del Artículo 29, en adelante las “Directrices” (Grupo de Trabajo Sobre Protección de Datos del Artículo 29, 2018). Este Grupo de Trabajo se creó en virtud del artículo 29 de la Directiva 95/94/CE. Se trata de un órgano consultivo independiente de la UE, en materia de protección de datos e intimidad. Estas directrices no son obligatorias, pero constituyen un sólido punto de referencia para la interpretación de las normas contenidas en el GDPR.

una clasificación o estimación general con aproximaciones de incerteza, pero difícilmente una decisión propiamente tal (Edwards & Veale, 2017, p. 46).

En tercer término, si bien a diferencia de la norma Europa, la chilena no incorpora como requisito que la decisión automatizada irroge efectos jurídicos o significativamente similares en los individuos, sí estimamos que esta referencia puede ser relevante al momento de aplicar la norma en Chile, si es que el Proyecto queda aprobado en los mismos términos. Según, las Directrices del Grupo del Artículo 29, la decisión debe afectar los derechos de una persona o comprometerlos de forma similar, por ejemplo: la cancelación de un contrato; la denegación de una prestación concedida por ley; la denegación automática de una solicitud de crédito en línea; decisiones que afecten el acceso de una persona a servicios sanitarios; decisiones que denieguen una oportunidad laboral a una persona o la sitúen en una posición de gran desventaja; decisiones que afecten el acceso de una persona a la educación; y decisiones que provoquen diferencias de precios sobre la base de datos o características personales; entre otros casos. Pero las Directrices también señalan que existe una decisión automatizada en la publicidad dirigida basada en la elaboración de perfiles<sup>7</sup>. En ese caso, la afectación al usuario se puede dar en las siguientes circunstancias: en el nivel de intrusismo del proceso de elaboración de perfiles (incluido el seguimiento de las personas en sitios webs, dispositivos y servicios); en las expectativas y deseos de las personas afectadas; en la forma en que se presenta el anuncio; y en el uso de conocimiento sobre las vulnerabilidades de los interesados (Grupo de Trabajo Sobre Protección de Datos del Artículo 29, 2018, p. 24). La interpretación del Grupo del Artículo 29 parece excesiva, especialmente, respecto de la publicidad dirigida en la elaboración de perfiles. Aplicando el criterio antes mencionado, casi todo comportamiento en línea estaría bajo sospecha. Supongamos el caso de los avisos que se despliegan en publicidad en línea para la compra de un producto determinado, por ejemplo, imaginemos que Alisa busca destinos turísticos en Portugal y como consecuencia de ello, le aparece publicidad en línea relacionada con paquetes turísticos para viajar a dicho destino. Cabe preguntarse, ¿hubo una “decisión” automatizada? No, la decisión la tiene finalmente Alisa. La decisión, ¿le produjo efectos jurídicos o similares a Alisa? No, ella solo vio el anuncio, no estando obligada a hacer clic en el mismo, de hecho, el aviso podría haber estado escondido en el evento que Alisa haya usado una herramienta de *ad blocker*.

---

7. El artículo 4 número 4) del GDPR define la elaboración de perfiles como: “*toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales de una persona física, en particular para analizar o predecir aspectos relativos al rendimiento profesional, situación económica, salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos de dicha persona física*”.

En cuarto término, en ambos artículos el responsable deberá adoptar medidas necesarias para asegurar los derechos del titular. Según las Directrices del Grupo del Artículo 29, dichas medidas deben incluir -como mínimo- un medio para que el interesado pueda obtener intervención humana, expresar su punto de vista e impugnar la decisión. Cuando se requiera la intervención humana, el revisor debe llevar a cabo una evaluación completa de todos los datos pertinentes, incluida información adicional facilitada por el interesado. La lógica de lo anterior descansa en que el interesado solo podrá impugnar la decisión automatizada o expresar su punto de vista si comprende plenamente cómo se ha tomado la decisión y sobre qué base (Grupo de Trabajo Sobre Protección de Datos del Artículo 29, 2018, p. 30). Junto con lo anterior, las Directrices sugieren una serie de medidas que pueden ser adoptadas por los responsables en el tratamiento de estos datos. Asumir que la intervención humana es deseable como medida necesaria para asegurar los derechos del interesado, genera igualmente una serie de inquietudes desde el punto de vista regulatorio. En primer lugar, supone que la intervención humana se encuentra exenta de sesgos, lo cual es falso. Todos los humanos tienen sesgos en mayor o menor medida, y éstos están dados por varios factores. En segundo lugar, depositar la confianza del sistema en la intervención humana, haciendo suponer que ésta es mejor, imparcial y más fiable que la IA basada en algoritmos de aprendizaje automático, puede desincentivar la adopción de estas tecnologías, partiendo de una premisa errada: que el humano ejecutará o revisará una tarea no solo de forma correcta, sino que más justa. Asumir lo anterior es tan errado como suponer que un humano puede efectuar mejores cálculos aritméticos que una calculadora. Por último, no explica cómo puede intervenir un humano en una decisión automatizada, cuando ésta no le es cognoscible ni a su diseñador, dado que forma parte de una operación algorítmica que se encuentra en una caja negra.

Otro problema que surge con las decisiones basadas en algoritmos y la pretensión de transparencia de éstos, consiste en que -en aras de dicha transparencia- se pueden vulnerar derechos de propiedad intelectual y secretos industriales de los titulares de los algoritmos. En efecto, los algoritmos son activos intangibles que le otorgan un gran valor económico a sus creadores o titulares, lo que se traduce una ventaja comparativa en el mercado. Para proteger esos algoritmos, los titulares pueden recurrir a distintas fórmulas, ya sea protegiéndolos como secreto industrial (por ejemplo: la fórmula de Coca Cola), como una obra (por ejemplo: el software propietario de Apple), o como una patente de invención (por ejemplo: Dropbox). La decisión final de cómo resguardar legalmente el algoritmo dependerá de su titular y de si concurren los requisitos legales para ser objeto de dicha protección. Sin embargo, en cualquier caso, la transparencia puede comprometer severamente la protección que el titular se esmeró en conseguir para su algoritmo, pues al revelar una decisión automatizada o la forma en que funciona un determinado algoritmo, se puede divulgar un secreto industrial,

ver cómo el algoritmo se encuentra escrito o cuáles son las principales características del mismo. Las soluciones normativas sobre este punto resultan insuficientes. El GDPR, en su Considerando 63 establece que el derecho de acceso contenido en ese cuerpo normativo “*no debe afectar negativamente a los derechos y libertades de terceros, incluidos los secretos comerciales o la propiedad intelectual y, en particular, los derechos de propiedad intelectual que protegen programas informáticos*”. El problema viene dado porque el Considerando 63 no es una norma vinculante, sino meramente referencial, lo que hace que la protección de los secretos comerciales y derechos de propiedad intelectual del titular del algoritmo, sean en la práctica, débiles.

El GDPR va a aún más lejos. Según las Directrices del Grupo del Artículo 29, la complejidad de un sistema no es excusa para entregar información al interesado. Fundamentan dicha afirmación en lo establecido en el Considerando 58 (igualmente referencial) que, en relación con el principio de transparencia, enfatiza su importancia pese a la proliferación de agentes y la complejidad tecnológica. Lo anterior no solo evidencia la contradicción interna entre los considerandos que componen el GDPR, sino la primacía de la transparencia por sobre la complejidad, ignorando la existencia de algoritmos en cajas negras. Y es porque finalmente, el GDPR no busca proteger a los creadores de tecnología e innovación, ni propende hacia un equilibrio normativo, sino que tiende exclusivamente a proteger a los usuarios de la tecnología y es allí, donde radica su principal debilidad: ver a la tecnología como una enemiga y a las empresas tecnológicas como una constante amenaza.

La inflexible y cerrada visión del GDPR grafica el problema de la aproximación normativa para resolver el tema de la transparencia. Un enfoque unilateral que se basa en un modelo *top-down*, en el que los gobernados por dicho modelo, no pueden intervenir en su configuración, sino solo, acatar so pena de elevadas multas, protegiendo así a los usuarios de la tecnología de la amenaza que supone ésta, en lugar de promover su uso.

Así las cosas, pareciera ser que el enfoque normativo proviene de la desconfianza que generan las grandes empresas tecnológicas y su creciente relevancia en la forma en que interactuamos. Sin embargo, las cortapisas que contiene dicho enfoque no afectan a esas empresas, que cuentan con el respaldo económico y técnico para cumplir cualquier requisito, restricción o imposición que establezca la autoridad de turno. Los verdaderos afectados, serán las pequeñas empresas y *startups* que luchan diariamente para conseguir financiamiento. El efecto final es evidente, las grandes empresas tecnológicas van a seguir en pie y las medianas y pequeñas quedarán en el camino, desincentivando de esa forma el desarrollo de nuevas tecnologías que, al final del día, van a beneficiar a toda la población.

En mérito de lo expuesto, es posible concluir que el enfoque normativo resulta insuficiente para resolver la transparencia algorítmica. Por el contrario, tiende a em-

peorar el ecosistema del desarrollo tecnológico estableciendo barreras de entrada que inhiben la innovación. Para solucionar dicho enfoque, en la siguiente sección exploraremos y rescataremos algunos recursos normativos, pero aplicando una solución global y pragmática sobre la base de políticas públicas flexible y equilibradas.

### **Soluciones para el problema de la transparencia algorítmica: una mirada pragmática e integrada**

En mérito de lo expuesto, es posible señalar que la transparencia es un valor deseable para la sociedad, pues permite a sus miembros, entender cómo funciona un sistema dado, y de esa forma generar confianza en su uso. Sin perjuicio de ello, existen ciertos casos que dificultan la consecución de dicho fin cuando un sistema es altamente complejo. Un ejemplo que ilustra lo anterior es la opacidad que se da en modelos de aprendizaje automático.

Una fórmula para resolver lo anterior es la estudiada en la sección precedente, esto es, resolver el problema de la transparencia y de los sesgos algorítmicos, a través de normas legales que impongan requisitos y obligaciones a los desarrolladores de estas tecnologías, so pena de multas y otras sanciones por su inobservancia. No obstante, dicha visión resulta parcial, insuficiente e inhibe el desarrollo de innovaciones. De hecho, la industria tecnológica europea que se ha desarrollado a la luz de estas regulaciones ha sido nula en comparación con Estados Unidos.

Prueba irrefutable de lo anterior, es el estudio efectuado por la firma de capital de riesgo MMC, que demostró que 40% de las empresas europeas catalogadas como “compañías de inteligencia artificial” no usan realmente IA. Según la investigación, hacen uso de la expresión “inteligencia artificial” para atraer más interés e inversión por parte de capital de riesgo. De hecho, en relación con empresas europeas que usan seriamente tecnología IA, 1 de cada 12 empresas nuevas emplea la IA como parte de sus productos o servicios, en comparación con la relación de 1 a 50 de hace seis años (Igleasias, 2019).

Sin perjuicio de lo anterior, en el evento que autoridades busquen regular la materia, en la siguiente sección esbozaremos soluciones desde una perspectiva integrada, que busca superar la estrecha e insuficiente visión normativa para abordar la problemática en estudio.

#### *Participación interdisciplinaria*

Según observamos en la sección precedente, el criterio normativo excluye a las empresas tecnológicas en la elaboración de políticas públicas. El inconveniente de dicho enfoque unilateral es que ignora si dichas empresas disponen de las capacidades técnicas y económicas para cumplir los requerimientos legales, especialmente, las pequeñas y medianas empresas.

De igual manera, un enfoque unilateral que solo beneficie a los ciudadanos y establezca exigentes requisitos para los desarrolladores tecnológicos, va a disminuir los incentivos de éstos últimos para crear soluciones que vayan en beneficio de la sociedad, dado el temor de verse expuestos a todo tipo de sanciones, generando así, barreras de entrada para el ingreso de nuevos actores. De esa forma, Chile va a seguir siendo un país consumidor de tecnología, en lugar de creador de la misma.

Una fórmula para resolver lo anterior, es establecer mesas de colaboración público-privadas en las que participe el regulador y gremios que agrupen a empresas tecnológicas. La idea es generar instancias de diálogos en las que el regulador señale qué es lo que se quiere conseguir a través de una eventual regulación, y los gremios a su vez, señalen qué es lo que están actualmente haciendo, y entre ambos, logren determinar qué es posible regular y cómo regularlo, si es que fuese necesario.

La idea de una Mesa Público-Privada no es nueva en Chile. Un ejemplo reciente fue la Mesa Público-Privada de Finanzas Verdes, liderada por el Ministerio de Hacienda y apoyada por el Banco Interamericano de Desarrollo (BID) y la Iniciativa Financiera del Programa de Naciones Unidas (UNEPFI, por sus siglas en inglés). El propósito de dicha instancia fue lograr un protocolo verde que considere los contenidos y variables específicas asociadas a temáticas verdes, como la medición del riesgo financiero del cambio climático; una declaración de criterios generales para mejorar el enfoque del mercado financiero y oportunidades del cambio climático; y una hoja de ruta de finanzas climáticas en Chile desde 2020 hasta 2024 (Ministerio de Hacienda de Chile, 2019).

### *Evitar fijar una definición legal de estándares de transparencia*

La transparencia en el ámbito tecnológico es un concepto dinámico, es decir, se encuentra en constante evolución. Por ende, los estándares que hoy se definen para dicha transparencia, el día de mañana podrían fácilmente quedar obsoletos, pues la velocidad con que la tecnología avanza es inmensamente superior a las normas que la regulan.

¿Significa lo anterior que no debe existir una regulación legal sobre la transparencia? No, pues ya existe el marco normativo para ello. En caso que la decisión sea basada únicamente en algoritmos complejos, el afectado podría -en caso de aprobarse el Proyecto- hacer uso del derecho de oposición a dicho tratamiento contenido en esa norma. En caso que la decisión sea adoptada por un humano que se apoya en algoritmos complejos, la responsabilidad recae en la persona que adoptó esa decisión y que puede explicarla al afectado, quien podría acogerse al principio de no discriminación arbitraria, en caso que dicha decisión contenga sesgos evidentes.

El peligro de fijar estándares legales -en el ámbito tecnológico- es que la rapidez con éstos pueden quedar obsoletos, dado el desarrollo de nuevas tecnologías. Un ejemplo de lo anterior es la regulación existente para operación de drones en Chile. La DAN 91, sólo se aplica a drones pequeños sobre 750 gramos de peso, fabricados con polietileno expandido o material equivalente, destinados al uso privado o recreacional<sup>8</sup>. Pues bien, uno de los principales fabricantes de drones a nivel mundial -DJI- lanzó el “*Mavic Pro*”, un dron con especificaciones técnicas sofisticadas y cuyo peso es de 743 gramos, quedando de esa forma, excluido de la aplicación de la normativa chilena existente la que, a dichos efectos, quedó obsoleta.

### *Escalas de explicación*

Una pregunta que conviene formular al momento de abordar la transparencia algorítmica es ¿deseamos una explicación de cualquier decisión adoptada por un algoritmo complejo? La respuesta es no. Los modelos de aprendizaje automático no siempre generan un impacto en los individuos, pues las soluciones que proveen son de distinta naturaleza, dependiendo de las necesidades de sus usuarios. Así, no resulta lógico comparar el impacto que genera el resultado de un algoritmo empleado para el otorgamiento de créditos o los usados en sistemas de salud para asignar prioridades de atención médica, con un algoritmo empleado para encontrar pareja o para traducir automáticamente un documento.

En virtud de lo expuesto, ¿cabe hacer exigible los mismos requisitos de transparencia a algoritmos de aprendizaje automático que resuelven problemas de distinta naturaleza? Pareciera que no. Por lo mismo, una propuesta en ese sentido es que -en caso que existan estándares de explicación normativos- éstos se apliquen conforme una escala con niveles de explicación dependiendo de los efectos que produce el modelo de aprendizaje automático: si los efectos negativos de una falla en el sistema son menores, es probable que las explicaciones a cómo funciona el mismo sean menos relevantes que en aquellos casos en que una falla podría poner en riesgo la vida de personas. Del mismo modo, si los usuarios pueden evitar fácilmente estar sujetos a la toma de decisiones automatizada, puede existir una menor expectativa o necesidad de una comprensión profunda (Google: 11).

En ese sentido, el riesgo de exigir un mismo estándar de explicabilidad normativo para cualquier sistema de aprendizaje automático, es que puede generar un impacto negativo y constituir una barrera de entrada para pequeñas y medianas empresas que desarrollan soluciones basadas en dichas técnicas y que no generan efectos nocivos en sus usuarios. En caso de existir estándares de explicabilidad normativos, éstos deberían tener un enfoque práctico, es decir, que otorgue claridad a un usuario cuando éste realmente se vea afectado por la decisión de dicho algoritmo.

---

8. Capítulo A, punto 91.102 de la DAN 91.

*Medidas alternativas para proveer explicación en casos de complejidad de los algoritmos en cajas negras (black box) y/o compromiso de derechos de propiedad intelectual*

Tomando en consideración las prevenciones ya señaladas, conviene enfocarnos en cómo se podría llevar a cabo -en la práctica- la explicación de la decisión tomada exclusivamente por un algoritmo de aprendizaje automático, es decir, sin intervención alguna de un ser humano.

La literatura examinada propone cómo y qué información debe ser proporcionada cuando la decisión ha sido tomada exclusivamente por un sistema automatizado. Así, las Directrices del Grupo del Artículo 29, señalan que: “*En lugar de ofrecer una compleja explicación matemática sobre cómo funcionan los algoritmos o el aprendizaje automático, el responsable del tratamiento debe considerar la utilización de formas claras y exhaustivas de ofrecer información al interesado, por ejemplo: las categorías de datos que se han utilizado o se utilizarán en la elaboración de perfiles o el proceso de toma de decisiones; por qué estas categorías se consideran pertinentes; cómo se elaboran los perfiles utilizados en el proceso de decisiones automatizadas, incluidas las estadísticas utilizadas en el análisis; por qué este perfil es pertinente para el proceso de decisiones automatizadas; y cómo se utiliza para una decisión relativa al interesado*” (Grupo de Trabajo Sobre Protección de Datos del Artículo 29, 2018, p. 35). Conjugando lo señalado por el Grupo del Artículo 29 con lo prescrito en el artículo 12 (1) del GDPR, la información, además debe ser concisa, transparente, inteligible, de fácil acceso, con un lenguaje claro y sencillo. El nivel de detalle en cuanto al contenido de la información proporcionada y la forma en cómo ésta se facilita, es necesaria para que el afectado pueda ejercer sus derechos<sup>9</sup>. En otras palabras, la transparencia opera como un factor habilitante para el ejercicio de derechos. Sin esa transparencia, los derechos consagrados normativamente, no tendrían valor alguno, pues no podrían ejercerse (Kaminski, 2019, p. 213).

De lo anteriormente expuesto, pareciera ser que la transparencia no supone un mayor desafío. La doctrina nos ha indicado qué información debe proporcionarse y cómo esa información debe ser proporcionada. Sin embargo, la literatura estudiada no repone ninguna solución a dos problemas que ya hemos tratado: (a) decisiones basadas en algoritmos de aprendizaje automático que operan bajo una lógica de caja negra (*black box*); y (b) qué sucede cuando al revelar una decisión se comprometen derechos de propiedad intelectual o secretos comerciales.

---

9. Dicho de otra forma, para ejercer el derecho a rectificación, el afectado debe primero ver qué es lo que hay que rectificar. De igual forma, para ejercer el derecho a no ser discriminado, el afectado debe primero ver qué factores se usaron para tomar la decisión que compromete sus derechos.

Frente a los desafíos que plantean los problemas expresados, una propuesta consiste en proveer de un conjunto de alternativas para que las empresas puedan utilizar en los dos casos antes mencionados. La elección de estas alternativas, por parte de las empresas, dependerá del tamaño de la empresa y su capacidad técnica, pero en ambos casos, la elección de la medida deberá ser fundada.

Dentro de estas medidas alternativas podemos destacar las siguientes:

(a) establecer canales de *feedback* para que usuarios del sistema reporten en caso que éste no funcione óptimamente<sup>10</sup>. De igual manera, podrían establecerse programas de recompensa de errores (*bug bounty programs*) en que se incentiva a *hackers* para que busquen errores en el sistema, pagándoles u otorgándoles reconocimiento por cada reporte de errores que encuentren. A través de estas vías, se podría mejorar la performance del sistema, evitando que éste genere futuros sesgos;

(b) establecer medios internos dentro de la aplicación para resolver controversias suscitadas por una decisión;

(c) utilización de técnicas de anonimización en el contexto de la elaboración de perfiles o promover el uso de técnicas de privacidad diferencial;

(d) testear con grupos de *hackers* (internos o externos) los modelos de aprendizaje automático, previo a su lanzamiento. Por ejemplo, preparando el algoritmo con datos específicos que podrían dar origen a sesgos, y esperar si el *output* es el previsto;

(e) auditorías externas, en ellas se podrían exigir: documentación sobre el propósito del sistema de IA, su función y rendimiento previstos; información sobre la arquitectura del modelo, los conjuntos de datos utilizados en la formación y las pruebas; comprobaciones internas realizadas para garantizar que se ajusta a los objetivos con que se diseñó el sistema, y una revisión de los procesos de la organización establecidos para supervisar las operaciones (Google: 12).

Quizás una solución que podría resultar lógica en virtud de lo examinado es que, al momento de preparar el algoritmo no se utilicen datos que podrían derivar en un resultado que genere sesgos. Sin embargo, tal como señalamos, muchas veces los sesgos no son evidentes y se encuentran implicados en los datos con que se entrena el algoritmo. Descartarlos en la etapa de preparación podría afectar la representatividad del sistema, empeorando su performance, generando resultados ineficientes, imprecisos y con un bajo grado de sofisticación.

---

10. Si el usuario tiene más control sobre la interacción con el sistema, puede ayudarlo a ser menos sesgado. Al mismo tiempo, el sistema puede contribuir al hecho que la persona se dé cuenta de sus propios sesgos cognitivos. En este sentido, sería ideal que las redes sociales o los buscadores -cuando la persona elige un resultado- dieran una valoración de si escoge como la mayoría o es una *rara avis*. También se podría mostrar más de una opinión y diferentes puntos de vista. Y, de paso, ayudaría a entender mejor el contexto en que cada cual se mueve (Baeza-Yates & Peiró, 2019).

### *Establecer regulatory sandbox en materia de datos personales para soluciones que reporten un gran beneficio social*

Con ocasión de la pandemia provocada por el Coronavirus (COVID-19), han proliferado las iniciativas para dar trazabilidad a los casos activos de la enfermedad y así disminuir el riesgo de contagio en la población. De igual manera, muchos laboratorios se encuentran diseñando una vacuna para atacar al virus.

En ambos contextos, el uso y tratamiento de datos personales ha sido clave. Sin embargo, la normativa europea sobre tratamiento de datos personales es extremadamente rigurosa, y no permite su tratamiento con la flexibilidad requerida, incluso en el contexto de una pandemia global. Por lo mismo, muchos países han promovido la adopción de esquemas regulatorios flexibles como los “*regulatory sandbox*”.

Los *regulatory sandbox* son marcos regulatorios que facultan la creación de una “caja” (“*sandbox*”) que permite poner en marcha proyectos que no cuentan con la autorización oficial para ello ni cumplen (al menos no totalmente) con la normativa sectorial pertinente. El objetivo de estas “cajas” -en el contexto de datos personales- es minimizar la incertidumbre respecto a la normativa aplicable, pues permite a las empresas con modelos de negocios innovadores en fase de desarrollo, conocer y adecuarse a la regulación de forma gradual y anticipada, y al regulador, entender de mejor manera el funcionamiento de una nueva tecnología, sobre todo en el contexto de una pandemia global.

Mientras la tendencia regulatoria analizada en la sección II de este artículo, busca imponer restricciones más rigurosas al uso de datos personales, lo que proponemos en este punto es lo contrario: permitir el uso de datos personales, bajo ciertas condiciones, para el desarrollo de soluciones tecnológicas que tengan un impacto favorable para la sociedad en áreas como la salud o educación. No se trata de desregular el tratamiento de datos personales, sino promover un uso responsable, con un adecuado balance entre la protección de los derechos de los individuos y la innovación tecnológica.

No es la restricción o prohibición en el tratamiento de datos personales lo que va a mejorar los modelos de aprendizaje automático, sino todo lo contrario: los empeora. Mientras más datos se dispongan, se pueden recolectar datos de mejor calidad y así disminuir los sesgos. Esto también puede incidir en extraer mejores atributos (o rasgos), para que el algoritmo los tenga en cuenta. (Baeza-Yates & Peiró, 2019).

### **Conclusión**

Un punto de referencia inicial del presente trabajo fue el cerebro humano y su complejidad en dos fases: en la toma de decisiones y en las explicaciones a las decisiones tomadas. Sobre este último punto, es importante destacar que nosotros no vivimos pidiéndoles explicaciones a las personas por sus decisiones, sino que tendemos a con-

fiar en ellas. Sin embargo, no ocurre lo mismo con los sistemas de IA, y más particularmente, con el aprendizaje automático. Por alguna razón más relacionada con la ciencia ficción, las autoridades tienden a desconfiar de la tecnología, proponiendo intervenirla a través de su regulación.

Los problemas que surjan en el contexto del desarrollo tecnológico se van a resolver gracias a la misma tecnología. Un ejemplo de ello fue el problema de la piratería. Diversos *stakeholders* rasgaron vestiduras exigiendo sanciones más severas a quienes descargaban contenido ilícito desde Internet. Gremios de la industria del cine y de la música condenaban esta práctica. Se introdujeron reformas legales tendientes a incrementar las penas para los infractores. Jóvenes fueron enviados a la cárcel por esta práctica. Sin embargo, no fueron las regulaciones las que derrotaron a la piratería, sino que fueron nuevos modelos de negocios tecnológicos. La irrupción y masificación de Netflix y Spotify hicieron que la descarga de una película o una canción, se haya convertido en una práctica obsoleta en la actualidad.

La existencia de sesgos en las decisiones tomadas por algoritmos de aprendizaje automático es un problema. De igual manera que lo es en el caso de humanos que toman decisiones sesgadas. Creer que los humanos no tienen sesgos, es paradójicamente, el sesgo más grande que existe. El grave error de quienes proclaman una mayor regulación en materia de transparencia consiste en pensar que los algoritmos de aprendizaje automático tienen una intencionalidad discriminatoria en los sesgos contenidos en sus decisiones. Peor aún, el reproche de culpabilidad se ha trasladado a las empresas que proveen el servicio o producto usando esta tecnología, invistiéndolas de una presunción de mala fe que resulta inverosímil.

Asumir lo anterior equivale a creer que los algoritmos de aprendizaje automático son infalibles cuando en realidad, son simples operaciones aritméticas que procesan datos para generar distintos modelos. Si los datos tienen implicados sesgos en su origen (*input*), esos sesgos volverán a reaparecer en sus decisiones (*output*), y será así mientras ese sesgo originario en los datos no sea reeducado o corregido mediante métodos que enmienden el sesgo muestral de forma mucho más sencilla y eficiente que la corrección de sesgos en los humanos.

El aprendizaje automático automatiza procesos más complejos a través de algoritmos que se van ajustando a medida que los datos que sirven para entrenar el mismo van evolucionando, permitiéndole reconocer variaciones en muchísimos factores, con la capacidad para ir mejorando su rendimiento con el tiempo.

El enfoque normativo estudiado en la sección II de este artículo ignora no solo el sustento científico que subyace en el aprendizaje automático, sino que también una premisa básica: la transparencia no es un problema legal, sino un problema tecnológico.

Una visión exclusivamente normativa, puede dar lugar a pérdidas de rendimiento y obstáculos burocráticos innecesarios para el desarrollo de tecnologías con repercusiones socialmente beneficiosas (Casey et al., 2019, pp. 181-182). De igual manera, la imposición de requisitos normativos de transparencia equivaldría a una prohibición de mejora o a una invitación a las empresas a eludir dichas normas (Bathae, 2018, p. 929). En definitiva, el costo social de establecer exigencias legales bajo un enfoque *top-down*, sin colaboración de la industria tecnológica y con onerosas sanciones, va a terminar promoviendo un futuro en el que solo se desarrollarán sistemas totalmente básicos predecibles pero transparentes, es decir, muy parecido a las calculadoras o a los computadores ya existentes.

El enfoque normativo establece exigencias regulatorias que generan verdaderas trabas en el uso de datos personales, desconociendo en su totalidad, el verdadero funcionamiento de herramientas de aprendizaje automático. Es muestra de la pretensión ilusa de autoridades que creen que, regulando un problema, éste desaparecerá. La evidencia muestra lo contrario, las regulaciones muchas veces generan más problemas.

Para tomar una decisión, el cerebro humano necesita información previa. Siguiendo los lineamientos normativos, que obstaculizan el tratamiento de datos personales y pone cortapisas a las decisiones basadas en el tratamiento automatizado de estos datos, cabe preguntarse, ¿qué decisión podría adoptar un modelo de aprendizaje automático que interactúe con usuarios, si no puede acceder a datos personales? En otras palabras ¿cómo se puede tomar una decisión sin información previa?

El afán regulatorio en aras del resguardo de los usuarios, tiene sentido en el contexto de datos personales al comprometer eventualmente los derechos ciudadanos. Sin embargo, su sobrerregulación es un reflejo de la incomprensión de la tecnología por parte de las autoridades. Hoy día resulta absurdo pensar que alguien desconfíe de una calculadora para realizar una operación matemática, pues existe confianza en el resultado obtenido y en su tecnología. Esa confianza es la que debería extenderse a toda innovación tecnológica. La base sobre la que se debe cimentar esa confianza se construye sobre el fino equilibrio entre disponibilidad de datos y el resguardo de la privacidad de los usuarios a través de distintas técnicas y modelos.

Finalmente, quisiera agradecer todo el invaluable apoyo y los comentarios del profesor John Atkinson-Abutridy de la Facultad de Ingeniería y Ciencias de la Universidad Adolfo Ibáñez, que permitieron otorgar luz en la materia estudiada.

## Referencias

- Baeza-Yates, Ricardo & Karma Peiró (2019). *¿Es posible acabar con los sesgos de los algoritmos?* Recuperado de <https://users.dcc.uchile.cl/~rbaeza/bias/sesgos-algoritmos.html>.
- Bagli, Charles (2018). Facebook Vowed to End Discriminatory Housing Ads. Suit Says It Didn't. *O The New York Times*. Recuperado de <https://www.nytimes.com/2018/03/27/nyregion/facebook-housing-ads-discrimination-lawsuit.html>.
- Bathae, Yavar (2018). The Artificial Intelligence Black Box and the Failure of Intent and Causation. *Havard Journal of Law & Technology*, 890-938. Recuperado de <https://jolt.law.harvard.edu/assets/articlePDFs/v31/The-Artificial-Intelligence-Black-Box-and-the-Failure-of-Intent-and-Causation-Yavar-Bathae.pdf>.
- Booker, Cory, Ron Wyden & Yvette Clarke (2019). *Algorithmic Accountability Act of 2019*. Congress Gov. Recuperado de <https://www.congress.gov/bill/116th-congress/senate-bill/1108/text>.
- Butler, Laura (2019). Evidence reveals risk assessment algorithms show bias against Hispanic population. *Phys Org*. Recuperado de <https://phys.org/news/2019-08-evidence-reveals-algorithms-bias-hispanic.html>
- Casey, Brayan, Ashkon Farhangi & Rolando Vogl (2019). Rethinking Explainable Machines: The GDPR's Right to Explanation Debate and the Rise of Algorithmic Audits in Enterprise. *Berkeley Technology Law Journal*, 145-187. Recuperado de <https://ssrn.com/abstract=3143325>.
- Comision Desafios del Futuro, Ciencia, Tecnologia e Innovación del Senado (2019). *Inteligencia Artificial para Chile: la urgencia de desarrollar una estrategia* [Documento PDF]. Recuperado de [https://inria.cl/sites/default/files/2020-04/Propuesta%20Estrategia%20IA%20Chile\\_1.pdf](https://inria.cl/sites/default/files/2020-04/Propuesta%20Estrategia%20IA%20Chile_1.pdf).
- Edwards, Lilian & Michael Veale (2017). Slave to the Algorithm? Why a Right To An Explanation" is Probably Not The Remedy You Are Looking For. *Duke Law & Technology Review*, 18-84. Doi: <http://dx.doi.org/10.2139/ssrn.2972855>.
- Galeano, Isabella (2019). 10 Conceptos Básicos de Inteligencia Artificial. *The Techno-lawgist*, 20-21.
- Google. (s.f.). *Perspectives on Issues in AI Governance* [Documento PDF]. Recuperado de: <https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf>.
- Grupo de Trabajo Sobre Protección de Datos del Artículo 29. (2018). *Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679*.

- Heinemeier Hansson, Jamie (2019). *I applied for an Apple Card. What they offered was a sexist insult*. Fast Company, 11 de noviembre. Recuperado de <https://www.fastcompany.com/90429224/i-applied-for-an-apple-card-what-they-offered-was-a-sexist-insult>.
- Henley, Jon & Robert Booth (2020). *Welfare surveillance system violates human rights, Dutch court rules*. The Guardian, 5 de febrero. Recuperado de <https://www.theguardian.com/technology/2020/feb/05/welfare-surveillance-system-violates-human-rights-dutch-court-rules>.
- Iglesias, Alberto (2019). *Desmontando el mito: el 40% de las empresas europeas de inteligencia artificial no usa realmente la IA*. Business Insider, 7 de marzo. Recuperado de <https://www.businessinsider.es/40-empresas-europeas-inteligencia-artificial-no-usa-384970>.
- Kagan, Justice (2018). *Equal Credit Opportunity Act (ECOA)*. Investopedia, 28 de marzo. Recuperado de: <https://www.investopedia.com/terms/e/ecoa.asp>
- Kaminski, Margot (2019). The Right to Explanation, Explained. *Berkeley Technology Law Journal*, 190-217. Doi: <http://dx.doi.org/10.2139/ssrn.3196985>.
- Kearns, Michael & Aaron Roth (2019). *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. New York: Oxford University Press.
- Kobie, Nicole (2019). *The complicated truth about China's social credit system*. Wired, 7 de junio. Recuperado de: <https://www.wired.co.uk/article/china-social-credit-system-explained>.
- Ministerio de Ciencia, Tecnología, Conocimiento e Innovación (2020). *Ministerio de Ciencia, Tecnología, Conocimiento e Innovación*. Recuperado de <http://www.minciencia.gob.cl/politicalA>.
- Ministerio de Hacienda de Chile. (2019). *Ministro de Hacienda encabeza lanzamiento de Mesa público-privada de Finanzas Verdes*. Ministerio de Hacienda de Chile, 3 de julio. Recuperado de: <https://www.hacienda.cl/sala-de-prensa/noticias/historico/ministro-de-hacienda-encabeza.html>.
- Norvig, Peter & Stuart Russell (2004). *Inteligencia Artificial: Un Enfoque Moderno*. Madrid: Pearson Prentice Hall.
- Real Academia de la Lengua Española. (2020). *Diccionario de la Lengua Española*. Recuperado de <https://dle.rae.es/algorithmo>.
- Sandri, Piergiorgio (2019). Por qué el algoritmo también discrimina en el trabajo. *La Vanguardia*. Recuperado de <https://www.lavanguardia.com/economia/20190411/461579327563/empleo-mujer-trabajo-algoritmo-seleccion-personal.html>.

- Sangüesa, Ramon (2018). Inteligencia artificial y transparencia algorítmica: It's complicated. *BiD: textos universitarios de biblioteconomía i documentació*, 41. Doi: <http://dx.doi.org/10.1344/BiD2018.41.12>.
- Scott, Mark & Laurens Cerulus (2018). Europe's new data protection rules export privacy standards worldwide. *Politico*. Recuperado de <https://www.politico.eu/article/europe-data-protection-privacy-standards-gdpr-general-protection-data-regulation/>.
- Senado de Chile (2019). *Comisión Desafíos de Futuro propone estrategia de Inteligencia Artificial para Chile*. Senado. Recuperado de <https://www.senado.cl/comision-desafios-de-futuro-propone-estrategia-de-inteligencia/senado/2019-09-12/174942.html>.
- Shalev-Shwartz, Shai & Shai Ben-David (2014). *Understanding Machine Learning: From Theory to Algorithms*. New York: Cambridge University Press.
- Simmons, Dan (2019). *6 Countries with GDPR-like Data Privacy Laws*. *Comforte Insights*, 17 de enero. Recuperado de <https://insights.comforte.com/6-countries-with-gdpr-like-data-privacy-laws>.

### Sobre el autor

CARLOS ARAYA PAZ es Abogado. Licenciado en Ciencias Jurídicas y Sociales de la Facultad de Derecho de la Universidad Adolfo Ibáñez. Diplomado en Propiedad Intelectual de la Pontificia Universidad Católica de Chile. Abogado Director del Estudio Jurídico Magliona Abogados. Correo Electrónico: [carlosarayapaz@gmail.com](mailto:carlosarayapaz@gmail.com).   
<https://orcid.org/0000-0002-0420-8314>

## CUHSO

Fundada en 1984, la revista CUHSO es una de las publicaciones periódicas más antiguas en ciencias sociales y humanidades del sur de Chile. Con una periodicidad semestral, recibe todo el año trabajos inéditos de las distintas disciplinas de las ciencias sociales y las humanidades especializadas en el estudio y comprensión de la diversidad sociocultural, especialmente de las sociedades latinoamericanas y sus tensiones producto de la herencia colonial, la modernidad y la globalización. En este sentido, la revista valora tanto el rigor como la pluralidad teórica, epistemológica y metodológica de los trabajos.

### EDITOR

Matthias Gloël

### COORDINADORA EDITORIAL

Claudia Campos Letelier

### CORRECTOR DE ESTILO Y DISEÑADOR

Ediciones Silsag

### TRADUCTOR, CORRECTOR LENGUA INGLESA

Aurora Sambolin Santiago

### SITIO WEB

[cuhso.uct.cl](http://cuhso.uct.cl)

### E-MAIL

[cuhso@uct.cl](mailto:cuhso@uct.cl)

### LICENCIA DE ESTE ARTÍCULO

Creative Commons Atribución Compartir Igual 4.0 Internacional