

DOCTRINA

La transparencia y la no discriminación en el Estado de bienestar digital

Transparency and non-discrimination in the digital welfare State

Alberto Coddou Mc Manus  y Sebastián Smart Larraín 

Universidad Austral de Chile

RESUMEN El uso intensivo y masivo de tecnologías digitales en la gestión y provisión de servicios, asistencia o programas sociales ha planteado diversos desafíos en materia de derechos humanos. Este trabajo se enfoca en los desafíos que la opacidad y la discriminación algorítmicas plantean al denominado *Estado de bienestar digital*. Desde esta perspectiva, el artículo pretende ser un aporte teórico en la creciente discusión sobre la gobernanza digital y su potencial distópico, cuestión que puede ser de interés para los debates que se darán en Chile, en particular, con respecto a los derechos sociales en el marco del proceso constituyente y a una futura política de inteligencia artificial.

PALABRAS CLAVE Estado de bienestar digital, discriminación algorítmica, opacidad algorítmica.

ABSTRACT The massive and intensive use of digital technologies for the management and provision of social benefits, programmes or assistance has entailed several challenges for human rights standards. This work focuses on the challenges that algorithmic opacity and discrimination raise for the so-called *digital welfare State*. Within this approach, the article attempts to be a theoretical contribution to the increasing discussion about digital governance and its dystopian potential, which may be of interest to the debates that are taking place in Chile, in particular, regarding social rights within the current constituent process and a future artificial intelligence policy.

KEYWORDS Digital welfare State, algorithmic discrimination, algorithmic opacity.

Introducción

El Sistema Alerta Niñez es un sistema informático desarrollado, gestionado e implementado por la Subsecretaría de la Niñez del Ministerio de Desarrollo Social y Familia chileno y que tiene por objeto estimar y predecir, a partir de datos de diversas fuentes administrativas, el nivel de riesgo de niñas, niños y adolescentes de sufrir vulneraciones a sus derechos (Valderrama, 2020). Con esta información, el Estado puede intervenir de manera temprana, evitando mayores riesgos para una población vulnerable. Por su parte, el 26 de enero de 2021, la Contraloría General de la República (CGR) de Chile anunció un plan de control público inteligente, una *CGR impulsada por datos*. El contralor general, Jorge Bermúdez, explicó que tienen acceso a una base de más de 1.300 millones de registros disponibles, lo que les permitiría desarrollar una serie de auditorías, por ejemplo, al *Bono Clase Media*.¹

Estos ejemplos demuestran que es indudable que vivimos una época de gobernanza digital, es decir, una época marcada por la gestión y decisión de asuntos públicos a través de tecnologías digitales. Según el ex relator especial de las Naciones Unidas sobre extrema pobreza y derechos humanos, Philip Alston, estas formas digitalizadas de programas estatales de prestaciones sociales, que se denomina Estado de bienestar digital, se definen como un conjunto de «sistemas de asistencia y protección social [que] se basan cada vez más en datos y tecnologías digitales que se utilizan para automatizar, predecir, identificar, vigilar, detectar, singularizar y castigar» (Alston, 2019: 2). A nivel global, hay países que han iniciado procesos de decisiones judiciales a través de algoritmos con impactos desastrosos en materia de discriminación racial y de género, y algunos teóricos han señalado que es probable que gran parte de las decisiones en materia de políticas públicas a futuro sean realizadas por robots o algoritmos (Coglianese y Lehr, 2017: 1.147). Estos procesos han aumentado considerablemente las asimetrías de poder entre Estado y población, principalmente en materia de acceso a la información. Los Estados conocen más que nunca a su población y los ciudadanos tendemos a conocer cada vez menos cómo se toman las decisiones en materia de políticas públicas.²

Aunque la gobernanza digital implica una revolución en la administración de beneficios y sistemas de asistencia social en países de todo el mundo, usualmente considerada como parte de la agenda de modernización del Estado, en este artículo nos enfocaremos en analizar sus riesgos o aspectos débiles, un campo poco explorado desde una perspectiva del estudio crítico de los datos y los derechos humanos. En

1. «Una CGR impulsada por datos», Contraloría General de la República, 2021, disponible en <https://bit.ly/36LF21e>.

2. Véase la descripción que hace Foucault (1976) de los sistemas panópticos, en los que los vigilados «son vistos, pero no llegan a ver nunca».

parte, ello se debe a la opacidad y poca transparencia que existe en torno al diseño, procedimiento y aplicación de los algoritmos, y a una visión apolítica sobre el valor de los datos y sus potenciales impactos discriminatorios. La comunidad internacional ha reconocido que las violaciones a derechos humanos provenientes de decisiones públicas automatizadas abarcan todo el catálogo de derechos.³ En este artículo, sin embargo, nos enfocaremos en dos aspectos particulares: la opacidad algorítmica y su relación con el derecho a la información y a la no discriminación. La decisión de enfocarnos en políticas sociales y, particularmente, en los desafíos que plantean la discriminación y la opacidad, obedece a la intención de cubrir un aspecto poco debatido en la literatura académica sobre nuevas tecnologías, que ha focalizado sus preocupaciones en libertades civiles y seguridad.⁴ Desde esta perspectiva, este artículo pretende ser un aporte teórico en la creciente discusión sobre la gobernanza digital y su potencial distópico, cuestión que puede ser de interés para los debates que se darán en Chile, en particular, con respecto a los derechos sociales en el marco del proceso constituyente y a una eventual política de inteligencia artificial.

El trabajo está dividido en cuatro partes: en la primera, haremos una introducción al marco teórico desde el que abordaremos la complejidad de nuestro objeto de estudio (los conceptos de teoría crítica de los datos y el Estado de bienestar digital); luego, en la segunda parte se analiza el peligro que la opacidad de los algoritmos representa para las decisiones democráticas relativas a las prestaciones y asistencia sociales; en la tercera parte se analiza la potencial discriminación producida por decisiones automatizadas y las diferentes formas en que se articula; y finalmente, en la cuarta parte formulamos algunas preguntas sobre cómo el proceso constituyente y la potencial consagración de derechos sociales y económicos deben abordar los desafíos que surgen de la revolución tecnológica, además de realizar un análisis meramente ilustrativo de las preguntas que genera la política de inteligencia artificial chilena.

Teoría crítica de los datos y el Estado de bienestar digital

Esta primera sección ofrece a quien lee un par de conceptos desarrollados desde distintas perspectivas disciplinarias para tratar de comprender los problemas y oportu-

3. El 5 de julio de 2012, 85 países firmaron la resolución del Consejo de Derechos Humanos, liderada por Suecia, afirmando la propuesta de que los mismos derechos que las personas tienen en el mundo *offline* también deben protegerse en el mundo *online*. Dicha declaración ha sido rectificada en años posteriores, siendo la última del 2018. Para ello, véase Consejo de Derechos Humanos, «Promoción, protección y disfrute de los derechos humanos en internet», Consejo de Derechos Humanos, 2012, disponible en <https://bit.ly/3cK6tw4>.

4. Ver «Economic, social and cultural rights and the internet, Global information society watch», Association for Progressive Communications (APC) e International Development Research Centre (IDRC), 2016, disponible en <https://bit.ly/39WxLxe>.

nidades que se generan con los desarrollos tecnológicos. La idea de que vivimos en una era de *revolución tecnológica*, que se vislumbra en la importancia que han adquirido los procesos automatizados de decisión basados en la recolección y gestión de datos, o en procesos de aprendizaje de máquinas (*machine learning*) que adquieren cierta autonomía en la gestión de asuntos humanos, no ha pasado desapercibida para la reflexión teórica.⁵ En este contexto, se ha avanzado de diversas formas para comprender esta revolución y se han desarrollado marcos teóricos para ofrecer alguna narrativa que nos permita entender las implicancias, desafíos y potenciales regulaciones que podemos implementar para compatibilizar el uso de estas nuevas tecnologías y el respeto por principios que consideramos fundamentales, como aquellos que protegen los derechos humanos. Aquí, nos enfocaremos en dos conceptos que ofrecen los marcos teóricos que creemos cruciales para comprender la magnitud de los desafíos que plantea la revolución tecnológica para las decisiones públicas relativas al bienestar material de las personas.

Estado de bienestar digital

El concepto que abordaremos en este subapartado supone una breve mirada panorámica a la idea de Estado de bienestar, presente en el discurso público al menos desde fines de la Segunda Guerra Mundial, cuando diversas democracias europeas comenzaron a garantizar el acceso a prestaciones y servicios sociales a sus ciudadanos en un contexto de crecimiento económico bajo un modelo de *capitalismo de Estado* (Esping-Andersen, 1990: 2). En términos generales, la idea del Estado de bienestar supone la creación de instituciones estatales que, coordinadas entre sí, garantizan que toda la ciudadanía de un Estado pueda acceder a un mínimo de prestaciones y servicios sociales, ya sea que estos mínimos se articulen a través de derechos sociales, de una red de protección social o una suerte de ciudadanía social (Marshall y Bottomore, 1998). Si bien existen diversas variantes o modelos concretos de Estados de bienestar (Esping-Andersen, 1990: 9), el contexto latinoamericano ha estado marcado por una continua e interminable transición hacia la consolidación definitiva de garantías universales o por la existencia de iniciativas fragmentadas que dan forma a lo que algunos autores llaman como Estados de bienestar emergentes (Huber y Niedzwiecki, 2014). De acuerdo con la literatura dominante, no podemos hablar de un modelo consolidado de Estado de bienestar en América Latina, sobre todo considerando la disparidad de beneficios y servicios públicos o privados a los que acceden los trabajadores formales por sobre los trabajadores informales y la gran cantidad de programas de asistencia social que son focalizados para una gran masa de la población que no alcanza a subsistir por sus propios medios (Huber y Niedzwiecki, 2014: 797).

5. Ed Pilkington, «Digital dystopia: How algorithms punish the poor», *The Guardian*, 14 de octubre de 2019, disponible en <https://bit.ly/3FmDVUy>.

Con el desarrollo de nuevas tecnologías de recolección, gestión y análisis de datos, recientemente los Estados alrededor del mundo, independientemente del nivel de bienestar que garantizan a sus ciudadanos, han generado sistemas de protección y asistencia social basados en procesos crecientemente automatizados y, en algunos casos, adoptando técnicas de *machine learning* que permiten el funcionamiento autónomo de estos sistemas. Si bien la idea de un Estado de bienestar digital data de la década de 1960, la radicalidad de la revolución tecnológica reciente ha generado procesos de automatización impensados hasta hace algunos años.⁶ En concreto, esto ha permitido a diversos Estados mejorar su capacidad de gestión de las prestaciones a las que acceden los ciudadanos, sobre todo a través del cruce de bases de datos que hasta hace un tiempo funcionaban de manera fragmentada, acumulando cada vez mayor información sobre la conducta de los usuarios de los programas sociales, muchas veces basados en tarjetas de identidad biométricas.⁷

Una visión panorámica de los diversos dispositivos del Estado de bienestar digital nos lleva a distinguir diferentes implementaciones: para la verificación del cumplimiento de requisitos o condicionalidades para acceder a beneficios sociales, como en el caso del Registro Social de Hogares chileno; para la predicción de conductas fraudulentas en su obtención, como el programa cuestionado en el famoso caso SyRI en Países Bajos (Lazcoz y Castillo, 2020); para calificar o evaluar conductas acorde a un sistema de crédito social, como en el caso chino (Engelmann y otros, 2019); o, por último, para evaluar eventuales vulneraciones a grupos de especial protección que ameritan una intervención estatal, como en el caso del Sistema Alerta Niñez recientemente implementado en Chile (Valderrama, 2020). La implementación de estos dispositivos depende de la capacidad estatal de recolectar y analizar datos, cuestión que muchas veces motiva a los arreglos institucionales a externalizar algunos servicios a actores privados, muchas veces sin los debidos resguardos o cautelas requeridas para la protección de los intereses de las personas o sus derechos básicos (Muñoz, Neffé y Pérez, 2021).

Por otra parte, y como ha quedado demostrado en estudios sobre el impacto de la economía de plataformas, los actores que acumulan datos suelen adquirir el poder de controlar o influir en la conducta de los actores de esas plataformas, generando diversas asimetrías de información que finalmente redundan en profundas desigual-

6. En la década de 1960 se observan los primeros intentos por relacionar programas que usan inteligencia artificial con decisiones legales (Ashley, 2017). Por ejemplo, en 1972, en Noruega se utilizaba un sistema de decisión completamente basado en decisiones algorítmicas para cuestiones relativas a vivienda y otros programas sociales (Schartum, 2016: 19). Dichas decisiones han sido objeto de críticas constantes por el potencial efecto en derechos humanos y la parcialidad con la que suelen resolver. Ver por ejemplo las críticas que se hacían en la década de 1970 (Sills, 1970; Tversky y Kahneman, 1974).

7. «Recognizing human rights norms in the “Principles on identification for sustainable development”: Civil society consultation report», Banco Mundial, 2020, disponible en <https://bit.ly/3txsAMe>.

dades de poder negociador entre las partes.⁸ Por ello, se ha avanzado en diversos enfoques críticos de la economía de plataformas, basados en los peligros que genera la acumulación y concentración de datos en pocos actores, como se ha destacado en el derecho de la libre competencia o en el derecho del consumidor.⁹ Sin embargo, cuando hablamos de la acumulación de datos en manos de programas sociales del Estado, usualmente la legitimación de mayor eficiencia y eficacia en el uso y gestión de recursos públicos parece justificar cualquier innovación tecnológica en la materia. Como recuerda el Banco Mundial (2017), el Estado se ha transformado, junto con las grandes empresas del *big tech*, en el principal productor (por ejemplo, miles de bases de datos que están disponibles de manera gratuita), consumidor (por ejemplo, gobiernos que compran bases de datos comerciales para verificar sus propias bases) y facilitador de datos (por ejemplo, a través de la regulación en el uso de datos), cuestión que ha despertado diversas preguntas en torno a la protección de la privacidad de los ciudadanos.¹⁰ A pesar de ello, las preguntas que genera el desarrollo incipiente de un Estado de bienestar digital van más allá de la privacidad, preguntas que creemos han estado relativamente ausentes del debate sobre los derechos humanos y la era digital.

En un contexto global marcado por la necesidad de competir con mercados a nivel mundial; con cadenas de producción global profundamente integradas e interdependientes; y en contextos domésticos marcados por una creciente austeridad fiscal y por la ausencia de poderes que hagan contrapeso al libre flujo de capitales a nivel mundial, la utilización de nuevas tecnologías en la gestión de programas sociales ha constituido la forma preferida de lograr, ahora sí, un uso eficiente de los recursos fiscales (Banco Mundial, 2017). Si a ello añadimos un discurso público y relaciones sociales marcadas por la ausencia de solidaridad, cuestión que fue clave en la legitimidad de los Estados de bienestar de la posguerra, la transformación digital de los programas de asistencia social se constituye en un régimen de control y disciplina de los más desaventajados, es decir, de quienes no pueden proveerse en el mercado los medios necesarios para su subsistencia (Laenen, Meuleman y van Oorschot, 2020). Como ya se está observando en los sistemas sociales de los Estados de bienestar de los países desarrollados, existe cada vez una mayor competencia entre diferentes intereses por recursos cada vez más escasos, que motiva a los gobiernos a buscar maneras

8. «World Employment and Social Outlook: The role of digital labour platforms in transforming the world of work», Organización Internacional del Trabajo, 2021, disponible en <https://bit.ly/3oGIMKn>.

9. Véase, por ejemplo, la «EU digital markets act» (2021) o las numerosas directivas de la Unión Europea sobre cuestiones contractuales o sobre derechos del consumidor: Directiva 93/13/EEC; 98/6/EC; 2005/29/EC; 2011/83/EU; y 2019/2.161.

10. «El derecho a la privacidad en la era digital», Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, 2018, disponible en <https://bit.ly/39Shirc>.

de mantener una apariencia de legitimidad, pero disminuyendo el costo fiscal (King, 2019). En este escenario, la automatización de sistemas para establecer la elegibilidad, el cálculo y el pago de prestaciones sociales; la masificación de sistemas digitales creados para detectar y prevenir el fraude a la asistencia social; el diseño de algoritmos para calificar a las personas en función de sus riesgos o necesidades específicas; o el aumento de plataformas en línea para interactuar y hacer un seguimiento de los beneficiarios de asistencia social, constituyen un pilar fundamental del Estado administrativo en la era digital (Gantchev, 2019).

Si bien el Estado administrativo digital fue desarrollado para simplificar la relación entre los ciudadanos y sus autoridades, las trabas u obstáculos burocráticos —como la necesidad de llenar formularios en línea o de comprender lenguajes crecientemente complejos— siguen tan presentes como antes.¹¹ Peor aún, como han documentado diversos expertos y estudiosos en la materia, la transformación digital de los programas y servicios sociales ha estado en la mira por las diversas violaciones a los derechos humanos que genera, en los más diversos contextos. A finales de 2019, el entonces relator especial de la ONU para la pobreza extrema y los derechos humanos Phillip Alston anunció que el mundo estaba «tropezando como un zombi en una distopía de bienestar digital».¹² En un ataque directo hacia los entusiastas de la digitalización y las promesas de un mejor acceso y transparencia, Alston señaló que, en asociación con el sector privado, los gobiernos están digitalizando el Estado del bienestar para «automatizar, predecir, identificar, vigilar y castigar a los más pobres», y llamó a una «reflexión seria sobre las desventajas» de la transformación de la protección y asistencia sociales. Como hemos visto en esta sección, si la discusión sobre lo que implica el Estado de bienestar supone cuestiones profundamente normativas, que apunta a los sentidos y ámbitos en que nuestras ideas de justicia material debieran tener expresión institucional, entonces la discusión sobre los desafíos del denominado Estado de bienestar digital no puede ignorar el estrecho vínculo que existe entre las dimensiones tecnológicas y las discusiones normativas (Alston, 2019: 5).

Los datos como forma de poder: Estudios críticos de los datos

El estudio crítico de los datos aporta una perspectiva particular para analizar la dimensión normativa de las tecnologías digitales, que surge desde un análisis crítico y pospositivista de las nuevas tecnologías (Boyd y Crawford, 2012; Kitchin, 2014; Kit-

11. Sophia Ranchordas, «Bureaucracy and vulnerability in the (digital) administrative state», *ICONnect Blog*, 2020, disponible en <https://bit.ly/2No1JHZ>.

12. «World stumbling zombie-like into a digital welfare dystopia, warns UN human rights expert», Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, 2019, disponible en <https://bit.ly/2N2o8V2>.

chin y Lauriault, 2014). Para complementar investigaciones puramente positivistas, en este artículo adoptamos la postura de que el estudio de datos necesita de una visión crítica para problematizar los supuestos de las decisiones tomadas en base no solo a grandes volúmenes de datos, sino también de datos que pueden cualitativamente reproducir sesgos; datos y decisiones que están inherentemente informados por ideologías que suelen permanecer ocultas en este tipo de estudios (Crawford, Miltner y Gray, 2014). Esto se vuelve especialmente relevante en un contexto en que los datos impregnan los discursos actuales en las ciencias naturales y sociales. Desde este punto de vista, creemos que su estudio crítico debe estar abierto a perspectivas culturales y éticas (Puschmann y Burgess, 2014).

El estudio crítico de los datos ha planteado preguntas sobre los marcos epistemológicos y los problemas sociales relacionados con los datos, imaginando distopías digitales plausibles o evaluando afectaciones reales a la vida e intereses de las personas. Esta perspectiva se transforma, entonces, en un intento formal de agrupar las investigaciones que cuestionan las formas despolitizadas que usualmente adopta la ciencia de datos, sobre todo a través del análisis de las formas en que los datos se generan o valorizan y cómo ejercen poder en las sociedades (Ananny y Crawford, 2018). En otras palabras, asumir una visión crítica de los datos significa reconocer que el *big data* siempre moldea y es moldeado por posiciones culturales contrapuestas (Boyd y Crawford, 2012) o que los procesos de *dataificación* —es decir el creciente registro de todos los aspectos de nuestras vidas en datos que pueden ser medibles, compartibles y transables— deben ser evaluados teniendo en cuenta las fuerzas estructurales que configuran la competencia capitalista (Morozov, 2019).

El término *estudios críticos de los datos* es usado inicialmente por Danah Boyd y Kate Crawford en su artículo «Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon», un texto que hoy se sitúa como un clásico de la literatura del área. En este artículo señalan que para analizar la gobernanza de datos se requiere situar la atención no solo en el programa o algoritmo generado, sino también en la posición política del gobierno que lo implementa y en los actores económicos involucrados en ellos, así como sus sistemas, procesos y servicios. Es decir, se deben entender los sistemas sociotécnicos detrás de un programa determinado, que se basa en el uso de la inteligencia artificial (Boyd y Crawford, 2012). Sin embargo, son Craig Dalton y Jim Thatcher quienes explícitamente acuñan el término, haciéndose la pregunta sobre qué significa un estudio crítico de los datos (Dalton y Thatcher, 2014). Ante esta pregunta, Rob Kitchin y Tracey Lauriault responden que los estudios críticos de datos deben enfocarse en la confluencia de elementos tecnológicos, políticos, sociales y económicos que enmarcan la generación, circulación y uso de datos (Kitchin y Lauriault, 2014). Tal como argumenta Kitchin, el uso de datos constituye, en clave de Foucault, una forma de conocimiento-poder;

una forma mediante la cual las personas, un fenómeno determinado o un territorio pueden ser vigilados y regulados (Kitchin, 2014).

Parece necesario, desde la teoría crítica de los datos, reflexionar sobre cuál ha sido el rol del derecho y de la regulación en la generación de una visión apolítica de los datos. Desde esta perspectiva, creemos que la teoría crítica de los datos debe dialogar con teorías político-económicas del derecho (Britton-Purdy, Singh Grewal, Kapczynski y Rahman, 2020). Por un lado, el derecho contribuye, a través de diversas herramientas, a configurar un terreno que impide o dificulta la intervención regulatoria del Estado y que facilita la acción privada. Así, por ejemplo, hoy la división entre lo público y lo privado, principalmente en un contexto digital, tiende a diluirse, generalmente en beneficio del privado y en desmedro de lo público. Bajo el lema de la eficiencia, las grandes plataformas tecnológicas privadas tienden a dominar áreas como educación, salud o transporte. Por otro, el derecho ha generado una especie de *lawlessness* de hecho. El derecho se ha transformado en un paso habilitante para sostener no solo la opacidad con que se actúa, sino también que sea el mercado quien controle la forma en que se recolectan, procesan y utilizan los datos (Zuboff, 2020). Desde una teoría crítica del derecho de los datos, se debería, por ejemplo, generar una resistencia a su mercantilización. Más que verlos como un producto básico que puede ser transado en el mercado o donde potencialmente un individuo tendría la propiedad sobre ellos, una visión crítica debiese analizar la dimensión relacional (Viljoen, 2020) o potenciar su rol emancipador y solidario (Morozov, 2019).

Si bien la perspectiva de los estudios críticos de los datos puede ser cuestionada por ser una *esponja* que absorbe toda perspectiva crítica sobre el tema (Dalton, Taylor y Thatcher, 2016), hay ciertos usos de la teoría que parecen relevantes para el análisis de la gobernanza digital que tratamos en este artículo. Así, por ejemplo, Gernot Rieder y Judith Simon (2016) se preguntan si la evidencia de los datos es suficiente para generar políticas públicas informadas y si permite una buena gobernanza. Concluyen que, más allá de centrarnos exclusivamente en las posibles consecuencias del fenómeno del *big data*, podemos obtener información adicional examinando sus raíces sociales y políticas, pero también sus técnicas y epistemologías. Siguiendo una línea similar, Gavin J.D. Smith (2020) se cuestiona cómo las políticas públicas de vigilancia basadas en algoritmos afectan el derecho a la ciudad y el derecho al *rostro*. Por su parte, tomando un marco teórico de los estudios críticos de los datos, Joanna Redden analiza las prácticas estatales canadienses que descansan en *big data*, generando un mapeo de las políticas públicas que basadas en datos e identificando cuáles son los principales efectos negativos identificados. En dicho estudio, Redden (2018) concluye que los riesgos son mayores y más variados de lo que se suele discutir. En todos estos casos, el estudio crítico de los datos permitió una mirada hacia la economía política de los datos (Britton-Purdy y otros, 2020) y los contextos que rodearon el diseño,

gestión o implementación de los sistemas, alejándose de las visiones tradicionales acerca de la neutralidad o imparcialidad con que suelen presentarse los dispositivos tecnológicos que los manejan, usualmente al amparo de agendas de gobernanza o transformación digital.

Si bien los debates sobre la gobernanza digital o las agendas de transformación digital suelen reducirse a foros técnicos, la mayor parte de las preocupaciones está orientada hacia cuestiones relativas al derecho a la información, a la libertad de expresión o a la privacidad. Con el fin de complementar la literatura de los estudios críticos de los datos desde una perspectiva de derechos humanos, en lo que viene nos enfocaremos en analizar dos cuestiones igualmente fundamentales: la opacidad en la toma de decisiones de bienestar social y económico y el potencial sesgo y discriminación que puede surgir de ella.

La opacidad en la toma de decisiones públicas automatizadas: Lo que (no) sabemos de las políticas sociales y económicas

Frente a la creciente opacidad en la toma de decisiones algorítmicas de políticas económicas y sociales, la solución de diferentes académicos y organizaciones de la sociedad civil ha sido hacer un llamado amplio a generar mecanismos de transparencia y rendición de cuentas de los algoritmos (Pasquale, 2015). La transparencia, en este contexto, es la afirmación de que alguien pueda mirar lo que hay detrás de tecnologías avanzadas de inteligencia artificial y gatillar formas de control de esos sistemas (Citron, 2008). La rendición de cuentas, por su parte, significa explicar, justificar y afrontar las consecuencias de los efectos negativos de los algoritmos (Boven, 2007). Además, existe un especial vínculo entre transparencia, rendición de cuentas y la idea de democracia. Para Danielle Citron (2008), de la existencia de reglas mínimas democráticas se derivan obligaciones como la de notificar y someter a consulta y comentario cualquier nueva decisión en materia de políticas públicas —por cierto, incluyendo políticas y procesos automatizados— y que la población tenga acceso a información básica sobre las formas de toma de decisiones. Otro aspecto fundamental de los estándares de transparencia y rendición de cuentas es que la autoridad se responsabilice de los impactos negativos de un proceso de toma de decisiones.

En términos generales, la rendición de cuentas supone «una relación entre un actor y un foro, en la que el actor tiene la obligación de explicar y justificar su conducta, el foro puede plantear preguntas y emitir juicios, y el actor puede enfrentar consecuencias» (Bovens, 2007: 450). Si bien esta definición surge teniendo en cuenta contextos de rendición de cuentas más tradicionales (como las relaciones político-burocráticas), esta conceptualización parece pertinente a la toma de decisiones algorítmicas. Por lo general, las decisiones algorítmicas en materia de políticas sociales y económicas son decisiones mixtas, donde un sistema de inteligencia artificial pro-

porciona información a un humano para que tome una decisión (Edwards y Veale, 2017). Ello no obsta la existencia de algoritmos completamente automáticos, donde un sistema de inteligencia artificial toma una decisión sin la participación humana. Sin embargo, esa forma de toma de decisiones es poco utilizada en sistemas públicos. En este sentido, la rendición de cuentas y la transparencia de un proceso determinado se ejercería, como en su versión clásica, respecto de la persona que basa sus decisiones de políticas públicas en la información obtenida de un algoritmo o en la persona que supervisa la decisión algorítmica (Burrell, 2016).

Existiendo cierto consenso en la literatura respecto de los problemas en materia de transparencia y rendición de cuentas de las decisiones algorítmicas, surge la interrogante de ¿por qué existe en la práctica esta caja negra? Esta pregunta es fundamental, principalmente porque la respuesta que demos guiará la forma de solucionar el problema de la opacidad. Jenna Burrell (2016) responde señalando que existirían al menos tres razones que explican la opacidad algorítmica. La primera razón sería la de una opacidad intencional por parte de empresas privadas que crean los algoritmos, a fin de generar autoprotección y encubrimiento institucional. La segunda potencial razón refiere al grado de opacidad derivado de la falta de capacidad generalizada para *escribir y leer* un algoritmo, una forma de opacidad inherente. La tercera, relacionada en parte con la anterior, tiene que ver con la complejidad de las decisiones algorítmicas que tienden a no ser compatibles con el razonamiento a escala humana; debido al alto número de datos analizados para tomar una decisión automatizada, estas no serían compatibles con las formas de razonamiento humano. A esta razón también se le llama opacidad inherente, donde ni los propios creadores del algoritmo son capaces de explicar sus resultados. Estos dos últimos argumentos son los más utilizados por académicos de las ciencias de la computación para explicar las dificultades prácticas de abrir los algoritmos o generar mecanismos de rendición de cuentas (Desai y Kroll, 2017).

Para enfrentar estos problemas, expertos como David Robinson y Alex Rosenblat discuten acerca de las diversas formas de regular los algoritmos.¹³ Más allá de las opciones institucionales que existen, ellos discuten críticamente la idea de justificar la opacidad algorítmica en base al desconocimiento o falta de capacidad técnica, argumentando que, independientemente de la complejidad de un algoritmo, siempre habrá un ser humano en algún lugar que tiene que decidir cómo va a funcionar o si sus elecciones son aceptables o no. Dado aquello, tiene que existir un impulso para que el regulador abra la puerta de esa sala donde se está decidiendo qué es lo que se espera de un algoritmo en particular. No importa cuán complejos sean los algoritmos, siempre habrá una pizarra donde se esté diseñando. El desafío, para los reguladores y para las personas que comparten este interés, es descubrir cómo ingresar a esa sala.

13. David Robinson y Alex Rosenblat, «Governing an algorithm in the wild», *Data & Society*, 2020, disponible en <https://bit.ly/3ry7eN1>.

La caja negra en la que se diseñan e implementan los algoritmos de todo tipo —incluidos aquellos que toman decisiones o aportan datos para decisiones político-económicas— genera impactos en materia de derechos humanos. En este punto es preciso resaltar que el catálogo de derechos humanos que ha sido históricamente reconocido para una realidad no necesariamente digital también debe ser protegido en el mundo *online*. En palabras del Consejo de Derechos Humanos, «los mismos derechos que tiene la gente en el mundo *offline* también deben estar protegidos en el mundo *online*».¹⁴ Sin embargo, como ya se adelantó, en este artículo nos enfocamos en los problemas asociados a la opacidad y la discriminación algorítmica, principalmente en materia de derechos sociales y económicos. Desde una perspectiva de derechos humanos, el llamado estaría focalizado en que los sistemas automatizados se diseñen e implementen con transparencia, incluyendo mecanismos de rendición de cuentas, y evaluando los potenciales efectos discriminatorios. El fin último sería evitar que las decisiones algorítmicas reemplacen o eludan la responsabilidad política de quienes están sujetos al escrutinio público y a la rendición de cuentas por decisiones defectuosas.

Así, a modo de ejemplo, en Nueva York el legislador local ha comenzado a buscar fórmulas de participación pública para probar los softwares de sistemas públicos y para abrir los códigos que faciliten las denominadas *auditorías algorítmicas*, de modo que las personas prueben cómo funciona un algoritmo.¹⁵ Alternativamente, algunos expertos han propuesto trabajar con algoritmos estándares, cuyos hiperparámetros sean públicamente accesibles, de modo que los resultados puedan variar de acuerdo con los datos con que se entrene a los algoritmos, pero sin dejar de conocer la forma en que se diseñó el modo de aprender de los modelos de *machine learning* (Mitchell, 2020). En otras palabras, esta propuesta apunta a garantizar la explicabilidad de los procesos automatizados de decisión en materia de políticas sociales. Recientemente, existen diversos intentos por estandarizar algunos (híper)parámetros de algoritmos, en el marco de la Organización Internacional para la Normalización (ISO), que garanticen un mínimo de transparencia y explicabilidad.¹⁶ A pesar de ello, estos modelos no son enteramente decodificables, pues si la transparencia pudiera garantizarse tan solo conociendo un *código*, no sería necesario entrenar a los algoritmos con datos.¹⁷ De ahí que todas las iniciativas por mayor transparencia algorítmica tienden a topar-

14. Ver «Promoción, protección y disfrute de los derechos humanos en internet», Consejo de Derechos Humanos, 2012, disponible en <https://bit.ly/3ryNR6t>.

15. Jim Dwyer, «Showing the algorithms behind New York city services», *The New York Times*, 2017, disponible en <https://nyti.ms/3aKomoK>.

16. «Information and communications technology. ISO/IEC JTC 1: Vision, mission and principles», Joint Technical Committee for Information Technology (JTC1), disponible en <https://bit.ly/3cpmdmE>.

17. Ron Schmelzer, «Towards a more transparent AI», *Forbes*, 2020, disponible en <https://bit.ly/3cliRRR>.

se, en algún punto del proceso, con un punto ciego: puedo conocer cómo el algoritmo aprendió a aprender, pero no puedo conocer, en último término, cómo es que ese primer aprendizaje produjo los resultados que produjo, especialmente en procesos de aprendizaje profundo basados en redes neuronales (*deep learning neural networks*).

Recientemente, el Berkman Klein Center for Internet & Society hizo una cartografía de 36 legislaciones, principios y estrategias de inteligencia artificial que cubren diversas zonas geográficas (Fjeld y otros, 2020). Entre los puntos analizados se encuentra precisamente la forma en que estos instrumentos tratan la transparencia algorítmica. Solo dos de los 36 documentos analizados no contenían cláusulas específicas respecto de la necesidad de transparencia y explicación.¹⁸ Así, por ejemplo, la estrategia de inteligencia artificial del gobierno alemán señala expresamente que las personas solo podrían determinar si una decisión automatizada es parcial o discriminatoria si pueden examinar la base (los criterios, los objetivos, la lógica) sobre la que se tomó la decisión.¹⁹ Si bien la transparencia algorítmica se debiese dar en todo el ciclo de la inteligencia artificial, es decir, desde su diseño, desarrollo o implementación, muchas de las legislaciones han visto la transparencia como una cuestión binaria, es decir, un algoritmo es o no es transparente. En lo que viene, veremos distintas técnicas jurídicas para asegurar la transparencia y entrar, como dice David Robinson, a esa sala oscura, a esa caja negra.²⁰

Para asegurar la transparencia en la toma de decisiones, distintas legislaciones han incorporado el derecho a la explicación.²¹ Si se ha tomado una decisión por medios automáticos, existe el derecho a saber si se están recolectando datos para ese propósito, a saber, cómo se tomó tal decisión, a que una persona intervenga y a disputar y/o impugnar cualquier decisión automatizada. En algunos casos, este derecho sirve para explicar la lógica, el significado previsto, las consecuencias y funcionalidad general de un sistema automatizado de toma de decisiones.²² La posibilidad de comprender el funcionamiento general de tales procesos se conoce como *interpretabilidad* y se le distingue de la *explicación*, que tiene un carácter local o aplicado a una decisión

18. Estos eran los casos de instrumentos redactados por el gobierno chino y por la embajada del Reino Unido en Ciudad de México.

19. «KI Strategie [Estrategia de inteligencia artificial]», Gobierno Federal de Alemania, 2020, disponible en <https://bit.ly/3aJNMps>.

20. Ver David Robinson y Alex Rosenblat, «Governing an algorithm in the wild», *Data & Society*, 2020, disponible en <https://bit.ly/3ry7eN1>.

21. Este es el caso de la estrategia de los Principios éticos para la inteligencia artificial de Latinoamérica (2020) que lo reconoce en su principio 11, disponible en <https://bit.ly/3cQrlAo>; o la estrategia de inteligencia artificial del Reino Unido (2018), disponible en <https://bit.ly/3oX35uh>; o la de Dubai (2019), disponible en <https://bit.ly/3l6DDtj>.

22. Véase, por ejemplo, la estrategia de inteligencia artificial del Reino Unido (2018), disponible en <https://bit.ly/3oX35uh>.

particular (Muñoz, Neffe y Pérez 2021: 31). El derecho a la explicación puede obligar a revelar decisiones específicas, es decir, la justificación, las razones y las circunstancias individuales de una decisión automatizada específica, sobre todo cuando causan un daño evidente.²³ Así, por ejemplo, la declaración de política de inteligencia artificial de la Comisión Europea conecta el derecho a la explicación con el principio de no discriminación, en el sentido de que una inteligencia artificial asequible y comprensible sería crucial para minimizar «el riesgo de sesgo o error» o los eventuales daños discriminatorios.²⁴

Una de las legislaciones que más ha desarrollado el derecho a la explicación es la legislación europea que en su Reglamento General de Protección de Datos se refiere explícitamente a este derecho.²⁵ De acuerdo con lo planteado por Sandra Wachter, Brent Mittelstadt y Chris Russell (2017), existirían cuatro problemas para implementar el derecho a una explicación que abra la caja negra en el contexto europeo. El análisis de estos cuatro problemas sirve para ver cómo se puede implementar dicho derecho en otras legislaciones. En primer lugar, la dificultad de encontrar sistemas jurídicos que aseguren el derecho de explicación como una obligación legalmente vinculante. Este sería el caso del Reglamento General de Protección de Datos que, si bien reconoce el derecho, no establece herramientas jurídicas para su exigibilidad.²⁶ En segundo lugar, incluso si fuera legalmente vinculante, el derecho solo se aplicaría en casos limitados (cuando una decisión negativa fue exclusivamente automatizada y tuvo efectos legales u otros efectos significativos similares). En tercer lugar, explicar la funcionalidad de algoritmos complejos, sus sistemas de toma de decisiones y su justificación en casos específicos, es un problema técnicamente desafiante y que no necesariamente aseguraría la entrega de información completa. Finalmente, la apertura de los algoritmos podría eventualmente colisionar con potenciales secretos comerciales o con el derecho a la privacidad (Wachter, Mittelstadt y Russell, 2017). Creemos, sin embargo, que el argumento respecto a la dificultad para entender los algoritmos incurre en un problema de circularidad argumentativa, pues justamente lo que se busca es que las personas que diseñaron el algoritmo o tomaron una decisión basada en estos, sean capaces de explicar o rendir cuentas, aunque reconocemos que puede llegar a ser imposible en algunos casos de opacidad inherente.

23. Este es el caso de la estrategia de Dubai (2019), disponible en <https://bit.ly/3l6DDtj>.

24. «Communication artificial intelligence for Europe. Shaping Europe's digital future», Comisión Europea, 2018, disponible en <https://bit.ly/3lfrUsi>.

25. Reconocido en su artículo número 71. Reglamento general de protección de datos (2016), disponible en <https://bit.ly/3rjB4I2>.

26. El argumento principal es que el artículo 22 (3) enumera los requisitos mínimos que deben cumplirse para que una toma de decisiones automatizada sea legal. Siempre que se cumplan estos requisitos, la toma de decisiones automatizada es legal y cumple con el Reglamento General de Protección de Datos.

Una forma en que los sistemas jurídicos han planteado el problema de la transparencia es justamente generando un vínculo con el derecho a la información. Así, por ejemplo, la estrategia de inteligencia artificial alemana declara expresamente que el derecho a la información va más allá del derecho a conocer aspectos técnicos del algoritmo o a saber si este utilizó o no datos personales, sino que se extiende a la obligación de que dicha información se redacte en un lenguaje sencillo y sea fácilmente accesible.²⁷ Otra alternativa es obligar a las autoridades a notificar a las personas en caso de que algoritmos estén tomando decisiones que las afecten, como establece la estrategia de inteligencia artificial de Reino Unido.²⁸ Si las personas no saben cuándo están sujetas a decisiones automatizadas, no tendrían la autonomía suficiente para decidir si quieren o no estar sujetas a dichas políticas. La Comisión Europea va un paso más allá y señala que la posibilidad de rechazar o salirse de programas basados en decisiones algorítmicas no es suficiente, sino que las personas también deben estar informadas sobre cómo solicitar rendición de cuentas por parte de un ser humano que debe garantizar que las decisiones de un sistema se puedan verificar o corregir.²⁹ Desde este punto de vista resulta fundamental que los Estados generen mecanismos de rendición de cuentas para respetar la transparencia pasiva y activa. De esa manera se puede hacer frente a los problemas no solo de acceso a la caja negra, sino también dotar de herramientas jurídicas para que las personas potencialmente afectadas por estas decisiones exijan información y, eventualmente, reparación.

Las diversas formas de discriminación algorítmica

La casi totalidad de los reportes, informes o documentos que buscan elaborar principios o estándares para ser aplicados en los procesos automatizados de decisión desarrollados con técnicas de inteligencia artificial mencionan al principio de igualdad y no discriminación como uno de carácter fundamental (Fjeld y otros, 2020). Esta mención, compartida unánimemente por todos quienes han reflexionado críticamente acerca de estos procesos, no suele distinguir las formas particulares de discriminación que se producen en el ámbito de los sistemas de asistencia y/o seguridad social. Para ello, muchos abogan por tener un completo acceso al modo en que estos procesos recolectan, analizan y gestionan datos a partir de diversas fuentes o conocer el modo en que funcionan los algoritmos, con el objeto de evaluar el impacto discriminatorio o los efectos sobre grupos desaventajados. Como analizamos en el

27. «KI Strategie [Estrategia de inteligencia artificial]», Gobierno Federal de Alemania, 2020, disponible en <https://bit.ly/3aJNMps>.

28. «AI in the UK: Ready, willing and able?», House of Lords Select Committee on Artificial Intelligence, Report of Session 2017-19, 2018, disponible en <https://bit.ly/3DrRxwZ>.

29. «Communication artificial intelligence for Europe. Shaping Europe's digital future», Comisión Europea, 2018, disponible en <https://bit.ly/3lfrU5i>.

apartado anterior, las barreras que existen para conocer el contenido y el modo de funcionamiento de estas verdaderas cajas negras (Pasquale, 2015) —muchas veces entregadas a actores privados protegidos por un cúmulo de instituciones propias del derecho privado (propiedad industrial, intelectual, secretos comerciales, acuerdos contractuales, etcétera)— plantean serias dificultades para la aplicación del principio de no discriminación en estos ámbitos (Kapczynski, 2020). Más aun, estos compromisos transversales desconocen el hecho de que el marco normativo que ofrecen los regímenes jurídicos de la antidiscriminación no suele incluir parámetros cuantitativos certeros que nos permitan decir cuándo un resultado generado por un proceso automatizado de decisión puede ser catalogado como discriminatorio (Rieke, Bogen y Robinson, 2018).

Las potenciales afectaciones al derecho a la igualdad y a la no discriminación se distinguen de acuerdo con las particulares formas en que los sistemas de asistencia y prestaciones sociales están diseñados, que pueden derivar, a su vez, de los sesgos, prejuicios o estereotipos de las personas que participan en su diseño o implementación o de la recolección y análisis de los datos que se proveen a los procesos automatizados de decisión. Como punto de partida, habría que señalar que los algoritmos no son neutros, así como tampoco lo son los datos que se usan para alimentarlos y las personas que los diseñan e implementan.

Usualmente, hablamos de que los algoritmos tienen un sesgo estadístico cuando la información que ellos recolectan y posteriormente procesan no es representativa de la realidad o, incluso, del objeto de interés de los propios algoritmos. En el caso de algoritmos que gestionan programas de asistencia social, y tal como señala el informe de Phillip Alston, a propósito de los procesos automatizados de gestión del riesgo y de calificación de necesidades, «determinar los derechos de las personas sobre la base de las predicciones elaboradas a partir del comportamiento de un grupo de población genera muchos errores» (2019: 12). Un ejemplo clásico en este sentido son los indicadores o criterios que utilizan estos sistemas para generar formas de alerta temprana de potenciales infractores o abusadores del sistema. Así, como relata el caso con el que abre el famoso libro de Virginia Eubanks (2018) sobre el tema, su familia fue catalogada como potencial infractora por el hecho de solicitar varios reembolsos por alto costo, al poco tiempo de haber contratado el seguro de salud Medicaid en Estados Unidos. Este criterio, utilizado por el algoritmo para detectar potenciales casos de fraudes, y basado en información existente, recolectada y gestionada por el propio sistema, puede generar profundas injusticias y exclusiones en personas que, por cuestiones fortuitas, sufran problemas de salud en cualquier momento. De no mediar un control humano, adecuado y efectivo a cada circunstancia o caso concreto, los algoritmos cometen errores con un impacto significativo en la vida de las personas.

Otro gran problema, en este sentido, se da por la poca diversidad o inclusividad de los equipos técnicos a cargo del diseño de los algoritmos o de traducir los datos

en los *inputs* del algoritmo. Así, como se ha destacado en diversas informaciones de las empresas del *big tech* que más participación tienen en la creación de procesos automatizados de decisión, existe una radical ausencia de personas pertenecientes a grupos desaventajados, quienes usualmente sufren por el impacto desproporcionado que tienen las medidas adoptadas a propósito de programas de asistencia social (West, Whittaker y Crawford, 2019). En estos casos, el sesgo no depende tanto de las cuestiones técnicas involucradas en el diseño y gestión del algoritmo o en problemas relacionados con los datos proporcionados al sistema, sino de los prejuicios, estereotipos o sesgos culturales de quienes participan en la creación de los algoritmos aplicados en programas sociales.

Una afectación frecuente al derecho a la igualdad y no discriminación en el acceso a derechos sociales se puede dar por el diseño del algoritmo que permite procesos de decisión automatizados porque en el diseño se consideran criterios que directa o indirectamente discriminan en contra de las personas usuarias de estos sistemas. El hecho de que los sistemas de asistencia social de los actuales Estados de bienestar observen cada vez más características alejadas de la universalidad y tiendan a crear condiciones de elegibilidad o requisitos de acceso a las prestaciones, supone una mayor utilización de indicadores o criterios que distinguen a las personas de acuerdo con su ingreso o situación socioeconómica. Estas últimas categorías, consideradas explícitamente como categorías protegidas por parte importante del derecho comparado de la antidiscriminación, debieran generar obligaciones al momento de diseñar el algoritmo, de modo que las distinciones o segregaciones no sean potencialmente discriminatorias (Fredman, 2010). Desde este punto de vista, se justifica un escrutinio estricto sobre algoritmos que utilicen estos criterios para asignar prestaciones u ofrecer servicios de asistencia social, con el objeto de evaluar si el uso de estos criterios se encuentra justificado por buenas razones que apuntan no solamente a la austeridad o eficiencia, sino a determinar las reales necesidades de las personas usuarias de esos sistemas o beneficiarias de estas prestaciones. Sin embargo, este tipo de escrutinios se ve usualmente confrontado con la justificación de la focalización o de las condicionalidades como fines legítimos que el Estado puede invocar para justificar el ejercicio de un derecho.

Este tipo de discriminación se articula, fundamentalmente, como una discriminación indirecta o por el impacto desproporcionado que tiene una medida o acción, en apariencia neutra, en un grupo determinado. En el marco del derecho de la antidiscriminación, estos casos suelen distinguirse por la utilización de criterios o indicadores (*proxys*) relacionados con la situación económica de las personas e, incluso, con otras categorías configuradas interseccionalmente por grupos —por ejemplo, mujeres, niños o personas con discapacidad— que sufren de formas de discriminación particularmente intensas dentro de la población de menores ingresos. Para evaluar los casos de discriminación algorítmica indirecta, antes que analizar la intención

de los responsables del algoritmo o de buscar comparadores que permitan probar un tratamiento desigual, es importante analizar estas medidas que, en apariencia neutra, generan un impacto o un resultado desproporcionado en personas de una categoría protegida, poniéndolas en situaciones de particular desventaja (Lundberg, 2019).

Otra forma de afectación supone considerar el modo en que ciertos diseños de los sistemas de asistencia o prestaciones sociales se aplican en un contexto social determinado, constituyendo casos ilustrativos de discriminación indirecta. Así, por ejemplo, un sistema cuyos contenidos contemplen los potenciales efectos de la exclusión de ciertas personas, cualesquiera sean las categorías protegidas en juego, puede generar problemas por el contexto social en el que se implementa. De este modo, un sistema que depende casi enteramente de la relación o vinculación con las personas a través de sistemas en línea puede excluir a parte importante de la población que no tiene acceso a internet o que, teniendo, no es capaz de utilizar adecuadamente sus canales para acceder a determinadas prestaciones o interactuar con el sistema en caso de existir algún problema. En otras palabras, esta forma de afectación no radica en el diseño del algoritmo, sino en las formas de interacción de las personas con los procesos automatizados de decisión (interfaz). En términos del derecho internacional de los derechos humanos, esto se podría evaluar desde el prisma de la obligación internacional que pesa sobre los Estados de garantizar el acceso efectivo, sin discriminación de ningún tipo, a las prestaciones que son objeto de los derechos sociales.

Consideradas como un todo, las formas concretas que han adoptado los sistemas de asistencia y prestaciones sociales en la actualidad, tanto en países desarrollados como en desarrollo, pueden configurar un caso predilecto de discriminación institucional. Este concepto, acuñado a propósito de un informe enfocado en investigar las causas de los problemas de violencia policial en contra de los jóvenes afrodescendientes en el Londres de los años 90, supone que la discriminación se encuentra enquistada en instituciones cuyos procedimientos y arreglos institucionales están configurados para excluir a ciertos grupos del acceso a bienes o derechos considerados como fundamentales.³⁰ Ahora bien, si es un hecho que los procesos automatizados de decisión relativos a la gestión del Estado de bienestar están llenos de sesgos, la pregunta es por qué los seguimos usando. Una de las razones que se han dado es que, aun estando llenos de sesgos que excluyen y discriminan a ciertos grupos, los algoritmos tienen un margen de error mucho menor del que se derivaría de entregar esa misma gestión a personas humanas. En otras palabras, como señala Ricardo Baeza-Yates, los algoritmos bien diseñados —a pesar de tener sesgos— son justos de acuerdo con los parámetros que se les han dado (2018). Así, por ejemplo, en el caso de un algoritmo (Correctional Offender Management Profiling for Alternative Sanctions, Compas)

30. «The Stephen Lawrence Inquiry», William Macpherson of Cluny, 1999, disponible en <https://bit.ly/3rtqds2>.

para decidir sobre el otorgamiento de fianzas en el marco de procesos judiciales, se probó que si bien el algoritmo discriminaba tanto como un humano con respecto a los afrodescendientes o hispanos al momento de predecir los casos de mayor riesgo de reincidencia de crimen, al menos era consistente y sistemático, aunque mostrara los mismos sesgos hacia las minorías que los jueces humanos.

Derechos sociales, nueva constitución y política de inteligencia artificial en Chile

Desde el pasado 15 de noviembre de 2019, Chile inició un proceso para cambiar su Constitución Política. Las causas que llevaron al proceso constituyente son varias y, probablemente, todas ellas tienen algún grado de relación con las desigualdades en el ejercicio y resguardo de derechos sociales.³¹ En paralelo, durante el segundo semestre de 2019, Chile comenzó a desarrollar su política de inteligencia artificial. Como hemos visto hasta acá, ante un eventual reconocimiento de derechos sociales y económicos en la Constitución, las decisiones que se tomen para su ejercicio y disfrute recaerán —cada vez más— en políticas basadas en decisiones automatizadas. Es por ello que creemos que estos procesos no debiesen verse de forma independiente. Por el contrario, el reconocimiento de derechos sociales en la nueva constitución debe informar y dialogar con el proceso de creación de una política de inteligencia artificial que asegure principios mínimos en materia de transparencia y no discriminación.

Reconocimiento de los derechos sociales en la nueva constitución

El actual proceso constituyente chileno, que terminará proponiendo un nuevo texto constitucional para Chile a mediados de 2022, a ser ratificado en un plebiscito de salida, ha estado marcado por una demanda transversal por el reconocimiento de derechos sociales. Como lo han señalado diversos comentaristas, los derechos sociales —entendidos como una demanda por una mayor presencia e intervención del Estado en la regulación y prestación de servicios y provisión de asistencia social— han estado en el origen de los malestares del actual modelo constitucional (Atria, Salgado y Wilenmann, 2020; Ruiz, 2020). Considerando este contexto, es esperable que una de las discusiones más importantes al interior de la Convención Constitucional encargada de proponer el nuevo texto constitucional sea la de la determinación del contenido u objeto de los derechos sociales (es decir, aquello a lo que el Estado está obligado o, por el contrario, a lo que las personas tienen derecho), la forma de financiar las prestaciones (es decir, la discusión de cuál será el rol del Estado y los

31. Véase, por ejemplo, la sección de contexto del informe que hizo la Oficina del Alto Comisionado para los Derechos Humanos (2019) sobre su misión a Chile (30 de octubre-22 de noviembre 2019), disponible en <https://bit.ly/3oyZOAC>.

privados en el financiamiento de las prestaciones materiales a ser garantizadas) y de sus formas de provisión (es decir, el espacio que corresponda al Estado y a los actores privados en las formas de entregar estas prestaciones a la ciudadanía). Todas estas discusiones, además, se dan en un escenario donde diversas voces llaman a partir estas discusiones desde la necesidad de avanzar hacia un Estado social de derecho (Figueroa y Jordán, 2020; Viera, 2014).

La pregunta por la relación entre los derechos sociales y el Estado de bienestar digital parece estar relativamente ausente en el marco del actual proceso constituyente. Como destacan Muñoz, Neffe y Pérez (2021: 33) sobre los derechos constitucionales:

Los que más riesgo de vulnerabilidad corren a la luz del uso de sistemas de tomas de decisiones automatizadas poco transparentes o no explicables, corresponden principalmente a los derechos de igualdad, privacidad y protección de datos, debido proceso y acceso a un juicio justo, seguridad, autonomía, así como, acceso a información y libertad de expresión [sin mencionar a los derechos sociales].

En contraste, aquí queremos analizar esta relación enfocándonos en las diferentes dimensiones y obligaciones a que da lugar el reconocimiento de los derechos sociales, incluyendo las relaciones con algunos de los derechos mencionados antes.

En particular, y en lo que nos interesa aquí, la discusión sobre las formas de provisión de los derechos sociales ha estado marcada por la posibilidad de admitir la participación de actores privados en el marco de regímenes públicos que impidan generar sistemas de provisión segregados por precio, acceso y calidad. A consecuencia de ello, existe una preocupación primordial por justificar constitucionalmente la necesidad que tiene el Estado de intervenir en el mercado para regular la forma de provisión definitiva de las prestaciones que realizan el contenido de los derechos sociales. En otras palabras, la discusión no radica tanto en la necesidad de que siempre sea el Estado el prestador directo, sino en las garantías institucionales, jurisdiccionales o de otro tipo, que permitan al Estado realizar la promesa original de los derechos sociales de desmercantilizar a los individuos, al menos en ciertos ámbitos, de las relaciones de mercado (Esping-Andersen, 1990: 35).

Como corolario, una de las cuestiones fundamentales en el debate constituyente será la relación entre los derechos sociales y la articulación de una nueva cláusula de igualdad y no discriminación. En efecto, como han mostrado diversos análisis sobre el tema en el derecho constitucional comparado, las cláusulas de igualdad constitucional que son entendidas en su dimensión sustantiva constituyen garantías fundamentales para avanzar hacia una comprensión adecuada de los derechos sociales (Coddou, 2019; Ronconi, 2018). Incluso en países que consideran a los derechos sociales como normas programáticas o como directivas de política pública en sus textos constitucionales, las cláusulas de igualdad han servido para recordarles a los Estados sus obligaciones de asegurar un acceso igualitario y no discriminatorio a los sistemas

de asistencia o seguridad social.³² Así, por ejemplo, la jurisprudencia constitucional colombiana, haciendo referencia a la cláusula constitucional de igualdad, justificó recortes de sueldo de funcionarios públicos escalonados que aplicaban disminuciones mayores a los más altos sueldos y estableciendo mínimos salariales que no se verían afectados (Landau, 2014: 279); del mismo modo, otros ejemplos de la jurisprudencia constitucional europea con posterioridad a la crisis financiera de 2007-2008 muestran cómo las cláusulas de igualdad se utilizaron para prevenir que los recortes presupuestarios se hicieran sin antes realizar un análisis del impacto que estas medidas podrían tener en los grupos más desaventajados de la población (O’Cinneide, 2014: 189). En otras palabras, solo a la luz de una interpretación robusta de las cláusulas de igualdad, que usualmente incluyen categorías protegidas que motivan al Estado ya sea a sospechar de su utilización en el diseño o implementación de ciertas medidas, así como a la necesidad de intervenir en la realidad para proteger a los grupos desaventajados, es posible acercarse a una comprensión adecuada de los ideales de solidaridad y comunidad a los que apuntan los derechos sociales.

En el marco de la estrecha relación que existe entre la noción de igual ciudadanía, que derivaría de una nueva interpretación sustantiva de la igualdad constitucional, y la consagración de los derechos sociales, es necesario considerar las variables que hemos analizado en secciones anteriores. Para ello, proponemos los siguientes puntos de análisis que creemos pueden aportar a una mirada más integral de los diversos desafíos que implica el debate sobre un nuevo modelo en materia de derechos sociales, en el contexto de la era de la revolución tecnológica.

Una primera cuestión fundamental para tener en cuenta es la discusión sobre el grado de universalidad en la cobertura de los sistemas de asistencia y protección social. En efecto, quizás una de las tendencias más problemáticas del desarrollo del Estado de bienestar en las últimas décadas es la creciente focalización de coberturas y la generación de requisitos y condicionalidades para acceder a los beneficios, servicios y prestaciones sociales (Watts y Fitzpatrick, 2019). De ahí que diversos comentaristas hablan de la actual importancia que tienen los argumentos de mérito en la distinción de los *pobres meritorios*, de aquellos que no cumplen con las condiciones o que no quieren proveerse en el mercado para su subsistencia (King, 2019: 306). Este tipo de discursos ha justificado, en el último tiempo, una serie de reformas a los servicios sociales, como los límites temporales máximos de recepción de beneficios (por ejemplo, Temporary Assistance for Needy Families en Estados Unidos), los límites totales en beneficios que pueden recibir los ciudadanos (por ejemplo, Households Benefits Cap en Reino Unido), o las sanciones por el no cumplimiento de ciertas condiciones, como el ausentarse o atrasarse a una entrevista de trabajo, a sesiones de capacitación

32. «Economic and social rights in the Courtroom», Equal Rights Trust, 2014, disponible en <https://bit.ly/2YPoygT>.

o entrenamiento, a una cita médica o el rechazo de una oferta de trabajo. Según Jeff King, tanto la extensión como la intensidad de las sanciones han aumentado debido a que los Estados pueden aparecer ante sus comunidades respetando el lugar fundamental que tienen los derechos sociales, y al mismo tiempo, adecuándose a las nociones de mérito y justicia arraigadas en parte importante de la población, sobre todo de las clases medias, que ven en los inmigrantes o en ciertos grupos de la población una tendencia a abusar de estos sistemas (King, 2019: 307). Para asegurar el cumplimiento de estas condicionalidades, los Estados han debido crear mecanismos administrativos de vigilancia, fiscalización y sanción de los incumplimientos, que han descansado crecientemente en la utilización de tecnologías digitales y en procesos automatizados de decisión (Adler, 2016; Eubanks, 2018). En este primer sentido, un modelo constitucional de un Estado social de derecho debiera tender hacia derechos sociales universales que entreguen menos espacio para una utilización nociva de las tecnologías digitales que sospechen, vigilen y castiguen a sus ciudadanos.

Una segunda cuestión para tener en cuenta es el aseguramiento o garantía del acceso igualitario y efectivo al goce de los derechos sociales, de modo que las tecnologías digitales o de inteligencia artificial que se usen en las prestaciones de los servicios sociales no impliquen barreras u obstáculos que afecten a los titulares de estos derechos, o que impacten desproporcionadamente a los grupos más desaventajados de la población. En este sentido, es importante recordar la importancia de una las obligaciones generales más importantes que existen sobre el modo de cumplir con las obligaciones relativas a los derechos sociales: la obligación de acceso no discriminatorio, que vincula incluso a aquellos países que no reconocen o garantizan derechos sociales en sus ordenamientos jurídicos. De ahí que el mencionado informe de Phillip Alston hable de la necesidad de «promover la igualdad digital» (2019: 16-18). En otras palabras, a pesar de que un Estado no reconozca derechos sociales con carácter vinculante, si entrega beneficios, asistencia o prestaciones sociales de manera digital, debe proveerlas de manera igualitaria «sin discriminación alguna por motivos de raza, color, sexo, idioma, religión, opinión política o de otra índole, origen nacional o social, posición económica, nacimiento o cualquier otra condición social».³³ Considerando esto, es importante tener en cuenta las posibles barreras de acceso que se puedan dar por una creciente digitalización de los servicios y prestaciones sociales, que asumen, además de la posesión de dispositivos tecnológicos y de conexión a internet, una alfabetización digital que sigue siendo una brecha incluso en los países desarrollados (EU Commission, 2020). Como señala el informe de Phillip Alston, «siempre debe haber una opción viable que no sea digital» para no generar barreras infranqueables, además de acompañar cualquier tipo de transformación digital con

33. Ver artículo 2.2, «Pacto Internacional de Derechos Económicos, Sociales y Culturales», Organización de Naciones Unidas, 1976, disponible en <https://bit.ly/3oZaau4>.

«programas orientados a promover e impartir las aptitudes digitales necesarias y garantizar un acceso razonable al equipo necesario y un acceso eficaz a internet» (2019: 17). Por último, como recomienda el mencionado informe, y «a fin de reducir el daño que causan las suposiciones incorrectas y las decisiones de diseño equivocadas, los sistemas de bienestar digital deben diseñarse junto con los usuarios a quienes van destinados y evaluarse de forma participativa» (2019: 18).

Una tercera cuestión a considerar en el debate constituyente, según la visión que hemos expuesto aquí, consiste en garantizar que las decisiones sobre el bienestar social de las personas no dependan entera, exclusiva o determinantemente de sistemas automatizados de decisión, ya sea que estén o no basados en técnicas de inteligencia artificial. Emulando lo señalado en el artículo 22 del Reglamento Europeo de Protección de Datos, se debe respetar el derecho de toda persona «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». El cómo se articule esta cuestión en el debate constituyente tendrá que ver, eso sí, con la discusión sobre la reciente reforma constitucional que incluyó el derecho constitucional de autodeterminación informativa o con alguna garantía de debido proceso ante los procedimientos administrativos que supone la consolidación de un Estado de bienestar, cuestiones que no desarrollaremos mayormente aquí.

Política de inteligencia artificial

Durante el primer semestre de 2019, el gobierno de Chile, en línea con los avances y tendencia en materia internacional (Fjeld y otros, 2020), se embarcó en la creación de una política de inteligencia artificial. Para ello, convocó a un grupo multidisciplinario de diez expertos y expertas, se generó un trabajo interministerial y un proceso de consulta ciudadana que constó de dos etapas. La primera consistió en la posibilidad de comentar el índice tentativo de la estrategia (donde participaron 1.300 personas) y la segunda convocó a mesas regionales donde participaron un total de 400 personas. Basado en las recomendaciones recabadas en dichas instancias, se redactó un borrador de política de inteligencia artificial que entre el 15 de diciembre de 2020 y el 27 de enero de 2021 se sometió a consulta pública.³⁴ Tomando en cuenta dicho borrador, en esta sección analizaremos las propuestas en materia de decisiones algorítmicas de programas sociales y económicos estatales y sus potenciales efectos desde el punto de vista de los principios de transparencia y de igualdad y no discriminación.

En primer término, si bien el borrador de política de inteligencia artificial reconoce la creciente importancia de las decisiones algorítmicas en materia de políticas

34. Para escribir este artículo se analizó el borrador de la Política nacional de inteligencia artificial sometido a consulta pública. Actualmente, la política está publicada y disponible en <https://bit.ly/3reCFyR>.

sociales estatales, poco dice sobre los estándares específicos que se deberían utilizar en estos casos. En particular, el borrador menciona expresamente que «diversas aplicaciones de inteligencia artificial están transformando los servicios que prestan los gobiernos, con mejores evaluaciones para programas sociales y mejores canales de atención ciudadanos» (p. 8).³⁵ Sin embargo, no hace ninguna referencia expresa al denominado Estado de bienestar digital ni desarrolla en detalle sus eventuales impactos negativos. Tampoco se refiere expresamente a derechos humanos. Solo lo hace en su eje número 3, que trata sobre *Ética, aspectos legales y regulatorios e impactos socioeconómicos*, al mencionar que existen riesgos en materia de derechos humanos, particularmente referidos a la dignidad, la privacidad, la libertad de expresión y la no discriminación arbitraria en el uso de la tecnología. Sin embargo, esta referencia se hace desde una visión restringida a lo ético y no a la exigibilidad de dichos derechos humanos, al mencionar que los análisis explorados en dicha sección «van más allá de la legalidad, y motivan un debate ético: qué usos son correctos o incorrectos para los individuos, comunidades o la sociedad en general» (p. 56).

A pesar de que uno de los principios del borrador es precisamente una inteligencia artificial con centro en las personas (p. 17), la falta de un marco basado en los derechos humanos llama la atención. Primero, por la restricción de los potenciales problemas asociados únicamente al uso de la tecnología, cuando los estándares internacionales mencionan expresamente que los potenciales efectos negativos en materia de derechos humanos se sitúan en su diseño, desarrollo e implementación. Además, llama la atención la falta de un foco particular en esta materia, toda vez que Chile participa del Comité Asesor Multisectorial de la inteligencia artificial en la Agenda de Cooperación Digital de la Secretaría General de la ONU, donde los estándares de derechos humanos son uno de los cinco pilares de dicho programa y que, por lo demás, el propio borrador cita (p. 9).

Más allá de esa importante omisión, desde una perspectiva del estudio crítico de los datos y las teorías del Estado de bienestar analizadas en este artículo, creemos importante detenerse a pensar críticamente en cómo el borrador trata cuestiones relativas a la acumulación de datos por entes públicos y privados, de acuerdo con las directrices generadas en materia de transparencia y derecho a la explicación, y en materia de derecho a la no discriminación.

En primer lugar, la forma de presentar los potenciales problemas en materia de acumulación y uso de datos es difusa. Se tiende a confundir cuestiones relativas a la privacidad y protección de los datos con la necesidad de mayor uso y acumulación por parte de organismos públicos y privados. Uno de los párrafos contenidos en el borrador es sugerente de dicha dicotomía: «Para fomentar el desarrollo de in-

35. «Política nacional de inteligencia artificial. Borrador consulta pública», Gobierno de Chile, 2020, disponible en <https://bit.ly/3nOZoRp>

teligencia artificial y potenciar su desempeño es necesario propiciar un ecosistema donde existan repositorios abiertos y de alta calidad, pero que también sean seguros y resguarden los derechos de las personas. Al mismo tiempo, se debe promover la confianza y generar las condiciones para propiciar que múltiples organizaciones o empresas compartan datos en favor de potenciar el sector completo, a través de mejores y más datos» (p. 36). Llama la atención que, a pesar de basarse en estándares de resguardo de datos, se haga un llamado a propiciar que organizaciones o empresas compartan datos. Eso lo hace extensivo a las agencias estatales, quienes deben «disponibilizar la mayor cantidad de datos posible, procurando que sea en forma oportuna (es decir, cuando se requieren), de alta frecuencia y periodicidad» (p. 37) y pone pocas limitantes al uso y almacenamiento de datos sensibles al señalar que se deben generar mecanismos administrativos y tecnológicos que aseguren su protección, lo que no significa en ningún caso la generación de repositorios cerrados o inaccesibles. Esta forma de tratar los datos como si fueran inocuos despierta serios problemas desde la teoría crítica de los datos que plantea justamente la posición de que ellos son moldeados y moldean posiciones culturales y políticas contrapuestas (Boyd y Crawford, 2012).

Un segundo aspecto crítico del borrador es precisamente la forma de tratar la transparencia y explicabilidad. En su objetivo 3.1.1., al tratar sobre la gobernanza de los sistemas de inteligencia artificial, se menciona expresamente que uno de los objetivos prioritarios será la de «establecer recomendaciones de transparencia algorítmica que puedan ser piloteadas en áreas de riesgo de vulneración de derechos fundamentales o que requieran auditorías y entender las decisiones automatizadas». Tomando en consideración los estándares internacionales mencionados en la segunda sección de este artículo, sería importante adelantarse a ciertos debates en este aspecto, por ejemplo, de la potencial regulación de un derecho a la explicación en el sistema jurídico chileno. Acá se debiese referir a estándares que puedan ser utilizados en todas las fases de diseño e implementación de las decisiones algorítmicas, enfatizando en aquellas decisiones que se toman en materia de políticas sociales y económicas. Adelantar, por ejemplo, el potencial derecho de las personas que, sujetas a una decisión algorítmica, debieran ser comunicadas *ex ante* de que un determinado algoritmo tomará una decisión de política pública, y otorgando la posibilidad de salirse del programa y de exigir la debida rendición de cuentas, ejerciendo su derecho a la autodeterminación informativa.³⁶

36. En el caso de Chile, la política de inteligencia artificial debe considerar los proyectos de ley que modifican la Ley 19.628 de 1999, Sobre Protección de la Vida Privada, y la Ley 19.223 de 1993, que Tipifica Figuras Penales Relativas a la Informática. Véase el Boletín 1.144-07, que «Regula la protección y el tratamiento de los datos personales y crea la Agencia de Protección de Datos Personales», refundido por el Boletín 11.092-07.

Un tercer aspecto a considerar en el borrador es la forma en que se trata la no discriminación. Siguiendo la línea de casi todas las políticas y principios en materia de inteligencia artificial, el borrador considera como uno de los cinco principios guía el de no discriminación, al establecer expresamente que «la inteligencia artificial no debe discriminar en forma arbitraria ni ser utilizada en perjuicio de ningún grupo» y poniendo foco particular en la importancia de generar una política con enfoque de género y como grupo de especial protección a niños, niñas y adolescentes (p. 18). En el caso de la potencial discriminación contra las mujeres, el borrador reconoce:

Muchos de los potenciales efectos negativos tienen su origen en sesgos existentes en los datos con que se entrenan los algoritmos y/o la composición de los equipos de desarrollo de software y hardware relacionado a la inteligencia artificial (p. 73).

Se hace cargo, precisamente, de uno de los aspectos mencionados en la tercera sección de este artículo, es decir, de la necesidad de incorporar a grupos que usualmente no participan en la creación de algoritmos. Además, uno de los aspectos positivos del borrador es la mención explícita de la idea de luchar contra la discriminación indirecta basada en el género (p. 73), que es una de las herramientas cruciales en el derecho comparado para evaluar el impacto o los resultados de decisiones algorítmicas (Zuiderveen, 2020). Sin embargo, no se genera un desarrollo particular que trate sobre cómo el derecho a la antidiscriminación reduce los potenciales sesgos y exclusiones de los procesos automatizados de decisión en políticas sociales o económicas. En otras palabras, no establece deberes públicos de realizar análisis preventivos del potencial impacto discriminatorio de las decisiones sobre políticas sociales o económicas basadas en procesos automatizados de inteligencia artificial, dejando a las víctimas la responsabilidad de iniciar los reclamos, ya sea ante cortes u otros órganos administrativos. Recientemente, en Estados Unidos y en Reino Unido se han establecido deberes públicos de considerar, a la par del análisis económico de los programas sociales, los potenciales impactos discriminatorios de las medidas administrativas que son plenamente aplicables a la transformación digital de este tipo de programas.³⁷ Esto llama la atención pues, por el contrario, sí se establecen deberes y mecanismos dirigidos principalmente a evitar las discriminaciones en plataformas de comercio electrónico (objetivo 3.3.1, p. 65). Entre otros, se identifica la necesidad de procurar el consentimiento informado, la utilización de algoritmos explicables y justos, y la generación de instancias de revisión y rendición de cuentas que debiesen ser explícitamente aplicables también a políticas sociales y económicas adoptadas por el Estado.

37. Véase por ejemplo Equality Act, section 149, Reino Unido, 2010, disponible en <https://bit.ly/3oBlTXT>; Memorandum for modernizing regulatory review, Presidential Documents, 2021, disponible en <https://bit.ly/326O35f>.

Conclusión

Considerando el auge del Estado de bienestar digital y tomando una perspectiva del análisis crítico de los datos, en este artículo nos hemos enfocado en dos de los múltiples efectos y problemas que pueden tener las decisiones públicas automatizadas. En particular, hemos visto cómo la opacidad y la potencial discriminación algorítmica puede generar problemas evidentes en la gestión o provisión de servicios, asistencias o programas sociales. Partiendo de la base de que gran parte de las decisiones sociales y económicas hoy tienen algún grado de automatización, las preocupaciones sobre la necesaria transparencia y fórmulas de promover la antidiscriminación, se vuelven cada vez más apremiantes.

En la primera parte introdujimos no solo el avance global de un Estado de bienestar digital, sino también cómo podemos analizar las decisiones en este contexto desde una visión crítica de los datos. En particular, observando que ni los datos ni las decisiones —con distintos grados de automatización— escapan de un potencial sesgo político, social o cultural. Desde este punto de vista, y reconociendo que los efectos en materia de derechos humanos son múltiples y variados, el foco de este artículo estuvo puesto en su segunda sección en entender la falta de transparencia algorítmica y proponer algunos ejemplos que permitirían el derecho a la explicación tanto *ex ante* como *ex post* de las decisiones de políticas públicas sociales y económicas. Pero la transparencia no es suficiente, pues aún con un derecho a la explicación, las decisiones automatizadas pueden generar discriminaciones. Por ello, en la tercera parte se analizan las diversas formas en que las decisiones automatizadas pueden discriminar, haciendo un énfasis en el llamado a la necesidad de reforzar la prohibición de discriminación indirecta. Finalmente, se analiza cómo este marco teórico podría ser utilizado para analizar el caso chileno, que no solo atraviesa por un período de cambio constitucional, sino que en paralelo ha generado una política de inteligencia artificial.

A medida que avanza el ejercicio de las decisiones algorítmicas, el problema de la opacidad se vuelve más crítico. En este sentido, es necesario que los reguladores actúen con prontitud, pues pareciera ser que, al aumentar la complejidad de las decisiones algorítmicas, es cada vez más difícil evaluarlas y más difícil la rendición de cuentas ante una eventual decisión discriminatoria, por consiguiente. Desde esta perspectiva, un enfoque puramente ético, como en el caso de la política de inteligencia artificial chilena, no es suficiente. Se requiere de políticas y regulaciones que, basadas en derechos humanos, permitan la exigibilidad de los derechos a la información y a la exigibilidad de una potencial reparación cuando los derechos han sido vulnerados. Si bien escapa del análisis propuesto en este artículo, a la par de los modelos éticos sobre la inteligencia artificial se han creado mecanismos que permiten, justamente, *ponerle dientes* a estos principios, generando en última instancia un modelo jurídico que permite no solo la consagración de los principios de transparencia y

no discriminación en materia de políticas sociales y económicas, sino su concreción en la práctica. Para ello, es necesario pensar en formas de cumplimiento, tales como auditorías, debida diligencia, deberes positivos, organizaciones encargadas de la protección de los ciudadanos tipo *ombudsmen*, entre otros.

Referencias

- ADLER, Matthew (2016). «A new Leviathan: Benefit sanctions in the twenty-first century». *Journal of Law and Society*, 43 (2): 195-227. DOI: [10.1111/j.1467-6478.2016.00749.x](https://doi.org/10.1111/j.1467-6478.2016.00749.x).
- ALSTON, Phillip (2019). *Estado de bienestar digital y derechos humanos. Informe del relator especial para extrema pobreza*, A/74/48037. Disponible en <https://bit.ly/3aDonxXht>.
- ANANNY, Mike y Kate Crawford (2018). «Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability». *New Media & Society*, 20 (3): 973-989. DOI: [10.1177/1461444816676645](https://doi.org/10.1177/1461444816676645).
- ASHLEY, Kevin (2017). *Artificial intelligence and legal analytics: New tools for law practice in the digital age*. Cambridge: Cambridge University Press. DOI: [10.1017/9781316761380](https://doi.org/10.1017/9781316761380).
- ATRIA, Fernando, Constanza Salgado y Javier Wilenmann (2020). *El proceso constituyente en 138 preguntas y respuestas*. Santiago: Lom.
- BAEZA-YATES, Ricardo (2018). «Bias on the Web». *Communications of the ACM*, 61 (6): 54-61. DOI: [10.1145/3209581](https://doi.org/10.1145/3209581).
- BANCO MUNDIAL (2017). *Big data in action for government: Big data innovation in public services, policy and engagement*. World Bank Group. Disponible en <https://bit.ly/2YU9BNP>.
- BOVENS, Mark (2007). «Analyzing and assessing accountability: A conceptual framework». *European Law Journal*, 13 (4): 447-468. DOI: [10.1111/j.1468-0386.2007.00378.x](https://doi.org/10.1111/j.1468-0386.2007.00378.x).
- BOYD, Danah y Kate Crawford (2012). «Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon». *Information Communication and Society*, 15 (5): 662-679. DOI: [10.1080/1369118X.2012.678878](https://doi.org/10.1080/1369118X.2012.678878).
- BRITTON-PURDY, Jedediah, David Singh Grewal, Amy Kapczynski y K. Sabeel Rahman (2020). «Building a law-and-political-economy framework: Beyond the twentieth-century synthesis». *The Yale Law Journal*, 129 (6): 1.784-1.835. Disponible en <https://bit.ly/3nc7pyg>.
- BURRELL, Jena (2016). «How the machine ‘thinks’: Understanding opacity in machine learning algorithms». *Big Data and Society*, 3 (1). DOI: [10.1177/2053951715622512](https://doi.org/10.1177/2053951715622512).
- CITRON, Danielle (2008). «Technological due process». *Washington University Law Review*, 85 (6): 1.249-1.313. Disponible en <https://bit.ly/3conoEt>.

- CODDOU MC MANUS, Alberto (2019). «La igualdad en la Constitución que queremos». En Jaime Bassa, Juan Carlos Ferrada y Cristián Viera (editores), *La Constitución que queremos* (157-194). Santiago: Lom.
- COGLIANESE, Cary y David Lehr (2017). «Regulating by robot: Administrative decision making in the machine-learning era». *Georgetown Law Journal*, 105 (5): 1.147-1.223. Disponible en <https://bit.ly/3Cu5SrF>.
- CRAWFORD, Kate, Mary L. Gray, Kate Miltner (2014). «Critiquing big data: Politics, ethics, epistemology». *International Journal of Communication*, 8: 1.663-1.672. Disponible en <https://bit.ly/3wZxDHp>.
- DALTON, Craig y Jim Thatcher (2014). «What does a critical data studies look like, and why do we care?», *Society + Space*. Disponible en <https://bit.ly/3CmoR7q>.
- DALTON, Craig M., Linnet Taylor y Jim Thatcher (2016). «Critical Data Studies: A dialog on data and space». *Big Data & Society*, 3 (1). DOI:10.1177/2053951716648346.
- DESAI, Deven R. y Joshua A. Kroll (2017). «Trust but verify: A guide to algorithms and the law». *Harvard Journal of Law & Technology*, 31 (1): 1-64. Disponible en <https://bit.ly/3nrJiM3>.
- EDWARDS, Lilian y Michael Veale (2017). «Slave to the algorithm? Why a right to explanation is probably not the remedy you are looking for». *Duke Law & Technology Review*, 16 (1): 18-34. Disponible en <https://bit.ly/3clK5Ys>.
- ENGELMANN, Severin, Mo Chen, Felix Fischer, Ching-yu Kao y Jens Grossklags (2019). «Clear sanctions, vague rewards: How China's social credit system currently defines 'good' and 'bad' behaviour». *Proceedings of the Conference on fairness, accountability, and transparency (FAT* '19) Association for Computing Machinery* (69-78), Nueva York. DOI: 10.1145/3287560.3287585.
- ESPING-ANDERSEN, Gøsta (1990). *The three worlds of welfare capitalism*. Princeton University Press.
- EU COMMISSION (2020). *Digital economy and society index (DESI). Human capital*. Disponible en <https://bit.ly/3jrkrnV>.
- EUBANKS, Virginia (2018). *Automating inequality: How high-tech tools profile, police and punish the poor*. Londres: San Martin's Press.
- FJELD, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy y Madhulika Sriku-mar (2020) «Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI», *Berkman Klein Center for Internet & Society*. Disponible en <https://bit.ly/3aK89TP>.
- FIGUEROA, Pamela y Tomás Jordán (eds.) (2020). *7 propuestas para la nueva Constitución de Chile*. Santiago: Editorial Usach.
- FOUCAULT, Michel (1976). *Vigilar y castigar: El nacimiento de la prisión*. Madrid: Siglo XXI.
- FREDMAN, Sandra (2010). «Positive duties and socio-economic disadvantage: Brin-

- ging disadvantage onto the equality agenda». *European Human Rights Law Review*, 290 (3).
- GANTCHEV, Valery (2019). «Data protection in the age of welfare conditionality: Respect for basic rights or a race to the bottom?». *European Journal of Social Security*, 21 (1): 3-22. DOI: [10.1177/1388262719838109](https://doi.org/10.1177/1388262719838109).
- HUBER, Evelyne y Sara Niedzwiecki (2014). «Emerging welfare states in Latin America and East Asia». En Stephan Leibfried, Evelyne Huber, Matthew Lange, Jonah D. Levy y John D. Stephens (editores), *The Oxford handbook of transformations of the state*. Oxford: Oxford University Press. DOI: [10.1093/oxfordhb/9780199691586.013.43](https://doi.org/10.1093/oxfordhb/9780199691586.013.43).
- KAPCZYNSKI, Amy (2020). «The law of informational capitalism». *Yale Law Journal* 129 (5): 1.276-1.599.
- KING, Jeff (2019). «The future of social rights: Social rights as capstone». En Amartya Sen y Katherine Young (editores), *The future of economic and social rights*. Cambridge: Cambridge University Press.
- KITCHIN, Rob (2014). «Big Data, new epistemologies and paradigm shifts», *Big Data & Society*, 1 (1). DOI: [10.1177/2053951714528481](https://doi.org/10.1177/2053951714528481).
- KITCHIN, Rob y Tracey P. Lauriault (2014). «Towards critical data studies: Charting and unpacking data assemblages and their work». En Josef Eckert, Andrew Shears y Jim Thatcher (editores), *Thinking big data in geography. New regimes, new research*. University of Nebraska Press.
- LANDAU, David (2014). «The promise of a minimum core approach: The colombian model for judicial review of austerity measures». En Aoife Nolan (editor), *Economic and social rights after the global financial crisis*. Cambridge: Cambridge University Press.
- LAZCOZ, Guillermo y José Castillo Parrilla (2020). «Valoración algorítmica ante los derechos humanos y el Reglamento General de Protección de Datos: el caso SyRI». *Revista Chilena de Derecho y Tecnología*, 9 (1): 207-225. DOI: [10.5354/0719-2584.2020.56843](https://doi.org/10.5354/0719-2584.2020.56843).
- LUNDBERG, Ema (2019). «Automated decision-making vs indirect discrimination: Solution or aggravation?». Independent written essay, master's program in Constitutional Law and Human Rights, Umea University. Disponible en <https://bit.ly/2N69gVm>.
- LAENEN, Tijs, Bart Meuleman y Wim van Oorschot (editores) (2020). *Welfare state legitimacy in times of crisis and austerity: Between continuity and change*. Londres: Edward Elgar.
- MARSHALL, Thomas H. y Thomas Bottomore (1998). *Ciudadanía y clase social*. Madrid: Alianza.
- MITCHELL, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji y Timnit Gebru (2019). «Model cards for model reporting». *Proceedings of the Conference on fairness, ac-*

- countability, and transparency (FAT* '19) Association for Computing Machinery (220–229), Nueva York. DOI: [10.1145/3287560.3287596](https://doi.org/10.1145/3287560.3287596).
- MOROZOV, Evgeny (2019). «Digital Socialism? The calculation debate in the age of big data». *New Left Review*, 116: 33–67. Disponible en <https://bit.ly/3zhKjOJ>.
- MUÑOZ, Catherine, Jeanna Neefe Matthews y Jorge Pérez (2021) «Sistemas de toma de decisiones automatizadas: ¿De qué hablamos cuando hablamos de transparencia y del derecho a una explicación?». *Bits de Ciencia*, 21: 27–36. Disponible en <https://bit.ly/3FsRfXi>.
- O'CONNOR, Colm (2014). «Austerity and the faded dream of a 'social Europe'». En Aoife Nolan (editor), *Economic and social rights after the global financial crisis*. Cambridge University Press. DOI: [10.1017/CBO9781107337954.010](https://doi.org/10.1017/CBO9781107337954.010).
- PASQUALE, Frank (2015). *The black box society: The secret algorithms that control money and information*. Boston: Harvard University.
- PUSCHMANN, Cornelius y Jean Burgess (2014). Big data, big questions: Metaphors of big data». *International Journal of Communication*, 8: 1.690–1.709. Disponible en <https://bit.ly/3kMjZ5z>.
- REDDEN, Joanna (2018). «Democratic governance in an age of datafication: Lessons from mapping government discourses and practices». *Big Data & Society*, 5 (2). DOI: [10.1177/2053951718809145](https://doi.org/10.1177/2053951718809145).
- RIEDER, Gernot y Judith Simon (2016). «Datatrust: Or, the political quest for numerical evidence and the epistemologies of big data». *Big Data & Society*, 3 (1). DOI: [10.1177/2053951716649398](https://doi.org/10.1177/2053951716649398).
- RIEKE, Aaron, Miranda Bogen y David Robinson (2018). «Public scrutiny of automated decisions: Early lessons and emerging methods. An Upturn and Omidyar Network report». Disponible en <https://bit.ly/3np3uxU>.
- ROBINSON, David y Alex Rosenblat (2020). «Governing an algorithm in the wild». *Data & Society*. Disponible en <https://bit.ly/3ry7eN1>.
- RONCONI, Liliana (2018). «Repensando el principio de igualdad: alcances de la igualdad real». *Isonomía*, 49. DOI: [10.5347/49.2018](https://doi.org/10.5347/49.2018).
- RUIZ, Carlos (2020). *Octubre chileno. La irrupción de un nuevo pueblo*. Santiago: Taurus.
- SCHARTUM, Dag W (2016). «Law and algorithms in the public domain». *Nordic Journal of Applied Ethics*, 1: 15–26. DOI: [10.5324/eip.v10i1.1973](https://doi.org/10.5324/eip.v10i1.1973).
- SILLS, Arthur. (1970). «Automated data processing and the issue of privacy». *Seton Hall Law Review*, 1.
- SMITH, Gavin J.D. (2020). «The politics of algorithmic governance in the black box city». *Big Data & Society*, 7 (2). DOI: [10.1177/2053951720933989](https://doi.org/10.1177/2053951720933989).
- TVERSKY, Amos y Daniel Kahneman (1974). «Judgment under uncertainty: Heuristics and biases». *Science*, 185 (4.157): 1124–1131. DOI: [10.1126/science.185.4157.1124](https://doi.org/10.1126/science.185.4157.1124).
- VIERA ÁLVAREZ, Cristián (2014). «Estado social como fórmula en la constitución chi-

- lena». *Revista de Derecho* (Universidad Católica del Norte), 21 (2): 453-482. DOI: [10.4067/So718-97532014000200014](https://doi.org/10.4067/So718-97532014000200014).
- VILJOEN, Salome (en prensa). «Democratic data: A relational theory for data governance». *Yale Law Journal*. Disponible en <https://bit.ly/3CX7gmL>.
- WACHTER, Sandra, Brent Mittelstadt y Luciano Floridi (2017). «Why a right to explanation of automated decision-making does not exist in the general data protection regulation». *International Data Privacy Law*, 7 (2): 76-99. DOI: [10.1093/idpl/ix005](https://doi.org/10.1093/idpl/ix005).
- WATTS, Beth y Suzanne Fitzpatrick (2018). *Welfare conditionality*. Londres: Routledge.
- WEST, Sarah M., Meredith Whittaker y Kate Crawford (2019). *Discriminating systems: Gender, race and power in AI*. AI Now Institute. Disponible en <https://bit.ly/2YVyNXX>.
- ZUBOFF, Shoshana (2020). *La era del capitalismo de la vigilancia*. Barcelona: Paidós.
- ZUIDERVEEN BORGESIU, Frederik J. (2019). «Strengthening legal protection against discrimination by algorithms and artificial intelligence». *The International Journal of Human Rights*, 24 (10): 1572-1593. DOI: [10.1080/13642987.2020.1743976](https://doi.org/10.1080/13642987.2020.1743976).

Sobre los autores

ALBERTO CODDOU MC MANUS es abogado. Licenciado en Ciencias Jurídicas y Sociales por la Universidad de Chile. Máster en Derecho por la Universidad de Nueva York y Doctor en Derecho por University College de Londres. Profesor auxiliar del Instituto de Derecho Público, Universidad Austral de Chile. Su correo electrónico es alberto.coddou@uach.cl.  <https://orcid.org/0000-0003-2041-2304>.

SEBASTIÁN SMART LARRAÍN es abogado. Licenciado en Derecho por la Universidad Católica de Chile. Doctor en Derechos Humanos y Estudios Latinoamericanos por la University College de Londres. Académico de la Facultad de Ciencias Jurídicas y Sociales de la Universidad Austral de Chile. Su correo electrónico es juan.smart@uach.cl.  <https://orcid.org/0000-0003-3924-702X>.

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

EDITOR GENERAL

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

SITIO WEB

rchdt.uchile.cl

CORREO ELECTRÓNICO

rchdt@derecho.uchile.cl

LICENCIA DE ESTE ARTÍCULO

Creative Commons Atribución Compartir Igual 4.0 Internacional



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