

Perspectivas de derechos humanos en la implementación de marcos éticos para la inteligencia artificial

Lara Gálvez, J. C. (2020). *Perspectivas de derechos humanos en la implementación de marcos éticos para la inteligencia artificial*. En C. Aguerre, (Ed.). *Inteligencia Artificial en América Latina y el Caribe. Ética, Gobernanza y Políticas*. Buenos Aires: CETyS Universidad de San Andrés.

J. Carlos Lara Gálvez *

Resumen

La promoción de marcos, directrices y principios éticos para la formulación e implementación de sistemas de inteligencia artificial ha provisto de un marco de referencia a estas expresiones de desarrollo tecnológico cuyo despliegue implica altos riesgos para intereses individuales y colectivos. Sin embargo, ha existido cierta crítica a la proliferación de directrices no obligatorias, frente a otros marcos normativos de vigencia global, tales como el derecho internacional de los derechos humanos. A su vez, existe limitada comprensión de la transposición de este esquema normativo a desarrollos tecnológicos de avanzada, particularmente sobre sus consecuencias prácticas, y sobre la capacidad de hacer exigible su cumplimiento. El artículo constituye un esfuerzo por conectar operativamente los principios éticos sobre inteligencia artificial con el marco normativo de los instrumentos del derecho internacional de los derechos humanos, especialmente a escala latinoamericana, y por sentar las bases para promover una implementación práctica de exigencias éticas que a la vez reconozca, respete y promueva derechos fundamentales.

* Abogado, Licenciado en Ciencias Jurídicas y Sociales por la Universidad de Chile. Master of Laws (LL.M., Law and Technology) por la Universidad de California, Berkeley. Director de Políticas Públicas en Derechos Digitales - América Latina. Puede ser contactado en: jclg@riseup.net

Palabras clave

inteligencia artificial
algoritmos
automatización
derechos humanos
responsabilidad

Introducción

La última década ha sido testigo de la proliferación de marcos, directrices y principios para una ética de la “inteligencia artificial” (IA). Centrados, en particular, en la formulación, programación o implementación de sistemas de IA, pero a menudo también dirigidos a la regulación por el Estado del uso de IA. La tendencia aparece, al menos en parte, como reacción al crecimiento en el uso de ciertas formas de IA por parte de distintas industrias y de aparatos estatales, con la consecuente documentación de impactos negativos derivados de la delegación de aspectos relevantes de la toma de decisiones con impacto sensible en la población o en parte de ella en sistemas automatizados.

Esta proliferación de marcos éticos es parte de un debate que no se restringe a disciplinas académicas, sino que se extiende de manera lógica a una discusión sobre la gobernanza de la tecnología algorítmica a nivel de estados, instituciones multilaterales, y por supuesto, empresas de tecnología. A su vez, esa multiplicidad de partes interesadas es expresión de la búsqueda de soluciones que implican aspectos legales, éticos y tecnológicos de ideal coexistencia para la operación adecuada de la tecnología.

En tal sentido, los referidos documentos han permitido entregar uno o varios marcos de referencia frente a posibilidades de desarrollo tecnológico que involucran riesgos altos. Sin embargo, se trata todavía de iniciativas de sistematización de directrices no obligatorias, de adopción ideal pero no imperativa, en contraste con otros marcos de acción con aristas deónticas (esto es, normas) con validez reconocida a nivel global, como es el caso del derecho internacional de los derechos humanos. Es decir, las iniciativas éticas así formuladas se yerguen, en general, separadas del marco normativo que mejor identifica las áreas de impacto negativo en el uso de tecnología. A su vez, parece existir una comprensión limitada de la transposición de un esquema normativo amplio a desarrollos tecnológicos específicos de avanzada, particularmente sobre sus consecuencias prácticas, y sobre la capacidad de hacer exigible su

cumplimiento. En síntesis, la coexistencia de documentación de principios para la inteligencia artificial y de marcos con expresa referencia al derecho internacional de los derechos humanos son expresiones políticas distintas para reaccionar frente a los riesgos asumidos o descubiertos de estas tecnologías.

¿Cómo se conectan los principios éticos sobre inteligencia artificial con el marco normativo de los instrumentos, especialmente a escala regional? ¿Cómo promover una implementación práctica de exigencias éticas que a la vez reconozca, respete y promueva derechos fundamentales? Además de identificar la existencia de marcos éticos y su vinculación con derechos humanos, ya sea en términos de prevención de daño o de promoción positiva, se hace necesario anclar obligaciones concretas sobre el diseño y la implementación de sistemas de decisiones automatizadas o asistidas por la automatización.

Este documento pretende abundar en la literatura sobre la rendición de cuentas de la decisión automatizada, basada en el uso del derecho internacional de los derechos humanos, del marco vigente que conocemos como sistema interamericano de derechos humanos (SIDH), en general, y de las obligaciones y derechos contenidos en la Convención Americana sobre Derechos Humanos (CADH) y la Carta de la Organización de Estados Americanos (COEA), en particular, con el propósito de llenar algunos de los vacíos en la aplicación de marcos éticos. La perspectiva adoptada es la atinente al contexto latinoamericano: países mayormente en vías de desarrollo, en los que la tecnología no es una parte significativa de la producción ni de los ingresos nacionales, con consecuentes desventajas frente a los países en los que no solamente se desarrollan las tecnologías de automatización, sino donde también se discuten sus marcos de gobernanza, incluidos los códigos éticos.

Atendida la pertenencia al SIDH, los estados latinoamericanos cuentan, de forma adicional a su propia capacidad

normativa interna, con una serie de herramientas vinculadas a sus deberes para prevenir infracciones a los derechos humanos, estableciendo mecanismos de resguardo, monitoreo y rendición de cuentas, y otorgando mecanismos de reclamo para sus ciudadanías. Además, asociados a los deberes de prevención, protección y remedio de los estados, existe una expectativa de respeto de los derechos humanos por parte de las empresas privadas. Esa expectativa sirve como punto de contacto para una participación activa de los estados, haciendo valer los distintos niveles de responsabilidad a lo largo de la cadena de elaboración y de implementación de herramientas para la automatización.

Solo la literatura reciente¹ ha buscado una más cercana relación entre la gobernanza de la automatización y el derecho internacional de los derechos humanos, más allá de la identificación de posibles derechos afectados. Sin embargo, así como las tecnologías a la que se refieren, esos esfuerzos provienen de países donde parece existir una mayor capacidad de influenciar el desarrollo tecnológico desde etapas tempranas. La búsqueda de operativización, para países como los latinoamericanos, requiere también del reconocimiento de condiciones de desigualdad de desarrollo.

La segunda parte de este documento resume parte de los hallazgos en la literatura sobre el impacto en derechos fundamentales de la inteligencia artificial, en tanto familia de tecnologías. Para ello, nos basaremos en investigación existente, analizada a su vez dentro del marco normativo provisto por el SIDH. A ello agregamos la presencia de aspectos políticos y económicos que son también parte de un análisis cabal desde el punto de vista del

¹ Véase, por ejemplo, McGregor et al. (2019), que ha forjado un marco de aplicación del derecho internacional de los derechos humanos para la rendición de cuentas para los algoritmos.

derecho internacional de los derechos humanos, incluidas sus dimensiones colectivas, vinculadas a su vez a otras perspectivas sobre la gobernanza de la tecnología. Para estos fines, usaremos “inteligencia artificial” de manera laxa, sin referencia a una forma específica de tecnologías, sino a una familia de ellas que dependen del procesamiento de datos para la formulación de predicciones, incluyendo así tecnologías mejor identificadas como tecnologías para la decisión algorítmica o automatizada.

La tercera parte examinará las propuestas de marcos, directrices y principios éticos para la inteligencia artificial, especialmente centrados en la transparencia y en la prevención de algunos de los riesgos, pero que pueden resultar insuficientes para la gama de derechos posiblemente afectados e identificados en la segunda parte. Esta es la circunstancia que hace necesaria la implementación a escala de los estados de sus obligaciones frente al derecho internacional de los derechos humanos, incluida la acción para la responsabilidad de las empresas. Sostenemos, a ese respecto, nuestro apoyo a la creación de una normativa internacional vinculante sobre empresas y derechos humanos.

En la cuarta parte propondremos la aplicación del SIDH como un marco normativo adecuado, desde el punto de vista sustantivo y procedimental, para la gobernanza en países latinoamericanos para la inteligencia artificial, avanzando hacia una operativización. Además de la capacidad de identificar ámbitos de daño, el SIDH permite un rango de acción en toda la cadena de desarrollo de tecnologías de automatización, incluida la decisión sobre su adopción o no adopción, en un contexto de políticas públicas auxiliadas por la automatización. Recomendamos, sobre este punto, marcos legales actualizados para hacer valer la responsabilidad tanto estatal como privada. Por cierto, habiendo atendido que los principales responsables de la adopción de medidas en tal sentido son los estados, se mantendrá escepticismo frente a la acción estatal para la regulación de las tecnologías.



2

Inteligencia artificial y el impacto en derechos fundamentales

2.1. “Inteligencia artificial”

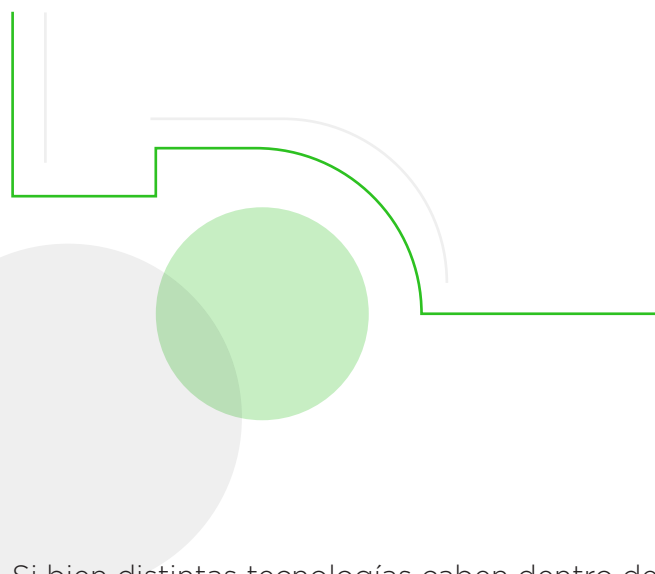
Existe vasta literatura desde hace al menos medio siglo que procura una definición para la inteligencia artificial. En términos muy generales, se trata de la tecnología que automáticamente detecta patrones en los datos, y hace predicciones a partir de ellos (Daly et al., 2019), simulando procesos cognitivos humanos, es decir, la “inteligencia”. Con menos atención en la tecnología, Muñoz León (2013)

se refiere a la IA como “aquel sistema que interactúa con su entorno seleccionando información mediante un determinado aparato epistémico, y que frente a la información considerada por el mismo sistema como relevante reacciona entregando respuestas conductuales consistentes”.

Las definiciones genéricas dan cuenta no solamente de la no referencia a una sola forma de aplicación del conocimiento mediante la técnica. Es decir que no hablamos de inteligencia artificial en referencia a una única tecnología, sino a una familia de tecnologías, que incluye técnicas tales como la visión computarizada, el procesamiento de lenguaje natural, los asistentes virtuales, la automatización de procesos robóticos, y el aprendizaje automatizado (Bughin et al., 2017).

Una característica común a las aplicaciones más habituales dentro de esta familia de tecnologías es el procesamiento de ingentes bases de datos, como forma de generar una mejor predicción, sea para clasificar o identificar objetos o para reproducir tareas complejas. Ese procesamiento de información aparece como una evolución de otros conceptos también reciente y ampliamente debatidos, como los de “decisiones algorítmicas”, “decisiones automatizadas”, o “Big Data”. El valor clave del uso de IA está en su poder de predicción, en tanto componente de la inteligencia (en términos generales, como característica asumida como humana), como forma de obtener información útil para la toma de decisiones, que al momento de la decisión no se tiene, sino que se concluye de datos existentes (Agrawal et al., 2018).

² Véase, por ejemplo, el informe de UNESCO, *Artificial Intelligence for Sustainable Development* (2019).



Si bien distintas tecnologías caben dentro de la noción de “inteligencia artificial”, el crecimiento en el interés en la academia y en la sociedad civil proviene fundamentalmente de su crecimiento en el mercado, como producto de un reportado aumento de usos de algunas formas de IA asociadas al crecimiento en la recolección y en el procesamiento de información. A la vez, se manifiesta cierto sentido de urgencia, especialmente en países en desarrollo como los de América Latina, para adoptar tecnologías a las que se atribuye un exagerado potencial de provisión de soluciones a una cantidad creciente de problemas (Elish y Boyd, 2017; Broussard, 2018; Citron y Jurecic, 2018), con una confianza en las máquinas más cercana al pensamiento mágico. Así, surge una aparente necesidad de aprovechar el potencial de estas tecnologías, con la correlativa preparación de los estados en la adopción de IA como probable factor de desarrollo en el actual escenario global² bajo el supuesto –no cuestionado– de que la IA es útil para el desarrollo.

En definitiva, la discusión sobre la gobernanza de la inteligencia artificial, en tanto familia de tecnologías, requiere una observación que no la aísle como objeto, sino que la contextualice como manifestación de desarrollo de industrias basadas en tecnologías digitales, tomando a su vez ese desarrollo dentro de un contexto más amplio de desarrollo económico.

2.2. El impacto de la IA en derechos humanos

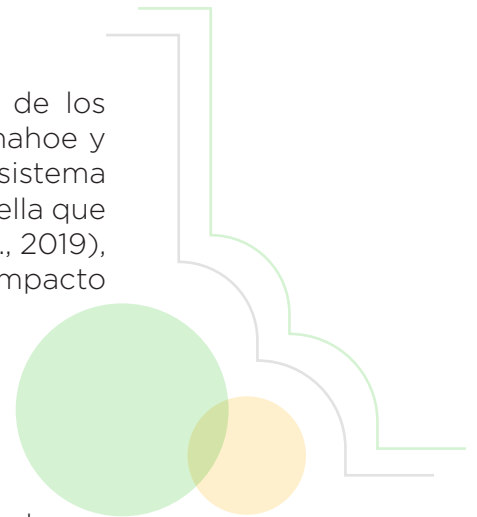
Existe un vasto cuerpo de literatura que identifica tanto las prácticas cuestionables de las empresas que utilizan datos a través de sistemas cada vez más avanzados para la consecución de sus fines, como los efectos negativos que ello conlleva. A medida que la técnica y el vocabulario del mercadeo varían, los riesgos del “Big Data”, del uso de “algoritmos” o de “inteligencia artificial” toman el protagonismo de la discusión. Y, en general, si bien las aplicaciones de algoritmos con diversos grados de complejidad se utilizan en aspectos cotidianos como la ordenación de contenidos o las recomendaciones en una red social, el impacto es mayormente acusado a propósito del uso de sistemas que para O’Neil (2016) constituyen armas de destrucción matemática en razón de su opacidad, de su escala y del daño que producen. Los ejemplos del uso negativo de sistemas de computación opacos en el sistema financiero (Pasquale, 2015), en la educación (O’Neil, 2016), en el sistema de persecución criminal (O’Neil, 2016; Angwin et al., 2016), en la entrega de beneficios sociales (Eubanks, 2018), y hasta en los resultados de búsqueda (Noble, 2018) han sido popularizados. Por cierto, los ejemplos siguen acrecentándose³. Una explicación indica que la naturaleza de las tecnologías, como constructo no neutral, implica la incorporación de las culturas y valores de quienes las diseñan y desarrollan, de lo cual no están exentas las tecnologías de inteligencia artificial, incluida la cultura de Silicon Valley (Broussard, 2018), en intensa yuxtaposición a los ámbitos culturales en los que pretenden hoy desplegarse estas tecnologías. Este punto es especialmente relevante para enfrentar los aspectos sistémicos de los impactos negativos, como discutiremos más adelante.

Un factor común a los hallazgos prácticos de situaciones de daño producido por algoritmos en general es la noción de una predicción, clasificación, o identificación o reconocimiento, que resulta errada o injusta. El consecuente efecto de las decisiones tomadas sobre la base de ese resultado (sea por un agente humano auxiliado por la tecnología o por un sistema programado para automatizar la decisión) constituye, de manera lógica, un impacto negativo en intereses relevantes (educación, seguridad social, justicia, empleo) concordantes con derechos fundamentales. Un cuerpo de literatura rápidamente creciente aborda esas

³ Al respecto, véase el trabajo del AI Now Institute, *AI Now Report (2017)*, disponible en: https://ainowinstitute.org/AL_Now_2017_Report.html, *AI Now Report (2018)*, disponible en: <https://ainowinstitute.org/litigatingalgorithms.html>, y *AI Now Report (2019)*, disponible en: https://ainowinstitute.org/AL_Now_2019_Report.pdf.

consideraciones desde la perspectiva del derecho internacional de los derechos humanos, variando desde consideraciones amplias (Donahoe y Metzger, 2019), pasando por una perspectiva analítica propia del sistema de derechos humanos (Wagner, 2018a; Raso et al., 2018), o por aquella que lo utiliza como marco para la rendición de cuentas (McGregor et al., 2019), hasta propuestas ambiciosas de marcos de evaluación de impacto (Mantelero, 2018).

Existe consenso en que la aplicación de sistemas para la decisión automatizada es capaz de impactar en un amplio rango de derechos. Wagner (2018a) conduce un análisis de los derechos más importantes afectados por las distintas aplicaciones de la familia de tecnologías de IA. Si bien su análisis se centra en la Convención Europea de Derechos Humanos (CEDH), el estudio apropiadamente destaca derechos que son también parte del SIDH.



Recogiendo el estudio descrito y contrastándolo con el rango de derechos reconocidos en el SIDH, podemos decir que el despliegue y la aplicación de sistemas algorítmicos y de inteligencia artificial, en particular en funciones públicas, puede afectar especialmente:

- el derecho a un juicio justo y a un debido proceso (artículo 8 CADH);
- el derecho a la vida privada (artículo 11.2 y 11.3 CADH), en particular en relación con la protección de datos personales;
- el derecho a la libertad de expresión (artículo 13 CADH);
- el derecho de reunión y la libertad de asociación (artículos 15 y 16 CADH);
- el derecho a un recurso efectivo (artículo 25 CADH);
- la prohibición de la discriminación y el derecho a la igualdad ante la ley (artículos 1.1 y 24 CADH);
- los derechos políticos vinculados a las elecciones (artículo 23 CADH); y
- los derechos económicos, sociales y culturales (artículo 26 CADH, en relación con la COEA).

En el estudio, Wagner (2018a) brinda ejemplos concretos en los que esos derechos se ven especialmente afectados dentro del contexto europeo. Sin embargo, desde las perspectivas de un posible análisis de riesgo y de un procedimiento de rendición de cuentas y responsabilidad, no resulta del todo útil medir una serie limitada de posibles derechos afectados.

Como contraste, el estudio de Raso et al. (2018) adopta una perspectiva basada en algunas aplicaciones destacadas de IA, y los posibles efectos que cada una de ellas puede tener en derechos humanos, incluidos posibles efectos positivos. Sin juzgar ni hacer propios los juicios sobre

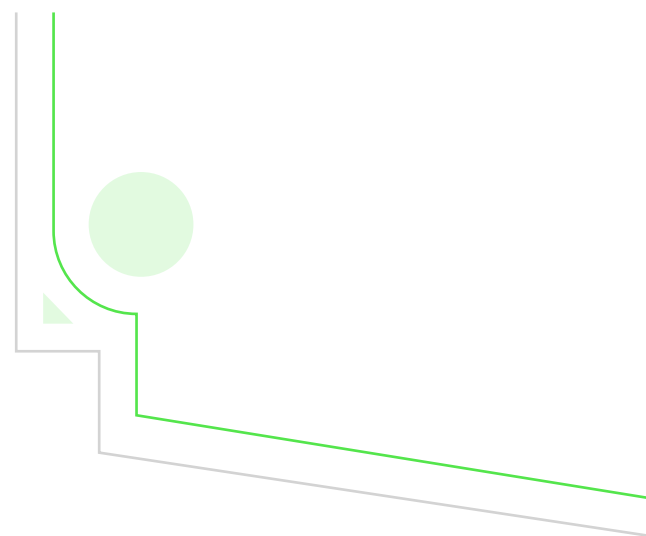
estos análisis, resulta particularmente valioso que el estudio no solamente dedique atención a formas extendidas de aplicación de IA como la calificación crediticia y la moderación de contenidos (también analizadas por Wagner [2019]), sino que agregue al análisis aspectos usualmente dejados de lado en un examen más profundo, como ocurre con los derechos económicos, sociales y culturales. El estudio de Raso et al. (2018), aun cuando toma una serie limitada de ejemplos, resulta útil para relevar la conveniencia de un análisis previo al desarrollo e implementación de IA, que en sus aspectos de consideración de derechos humanos cumpla con ser un proceso general, flexible y abierto, y a la vez informado por distintos ámbitos de pericia.



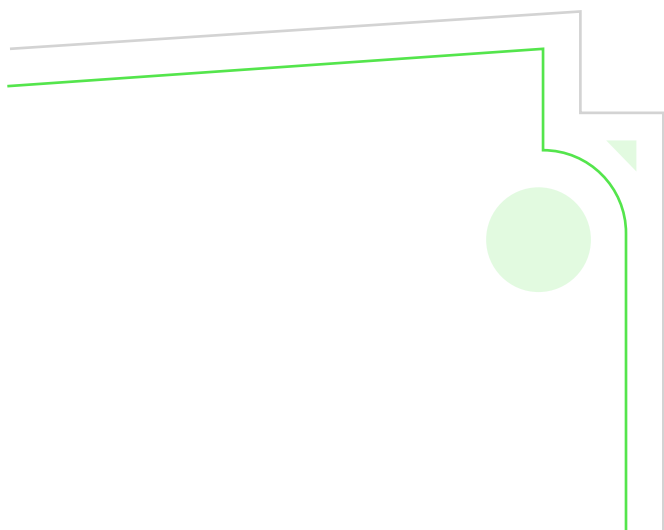
Por lo demás, un listado de derechos afectados por una tecnología o familia de tecnologías, en tanto desafío normativo, no es suficiente por sí solo para guiar la adopción de medidas de prevención y remedio frente a impactos nocivos. Un análisis de impactos en función de consideraciones normativas tiende a dejar de lado el estudio de las condiciones económicas, sociales y culturales, con aristas globales, que configuran la existencia y las dimensiones de esos riesgos. Se trata de una limitación que refiere tanto a los ámbitos de diseño y desarrollo de las tecnologías, como a los ámbitos de despliegue y de su aplicación. La discusión sobre la gobernanza de la inteligencia artificial, inclusive la centrada en derechos humanos, no solamente descuida las condiciones económicas de desarrollo de la tecnología bajo un sistema capitalista intrínsecamente desigual para quienes son objeto de aplicación de las tecnologías, sino que también corre el riesgo de ignorar las diferencias en los estadios de desarrollo entre las latitudes donde se concentra la producción tecnológica y donde pretenden desplegarse sus impactos. Una noción holística de los derechos humanos, que incluya aspectos de desarrollo, de derechos económicos, sociales, culturales y ambientales, y de las aristas colectivas de esos intereses, puede ser útil para economías en constante desventaja, como las de América Latina.

Resulta paradójico que buena parte de la discusión sobre ética y justicia en el uso de algoritmos de IA se centre en nociones de igualdad y rechazo a la discriminación arbitraria realizada o reforzada por tecnologías, a nivel individual, cuando existen dimensiones colectivas de desigualdad entre el mundo desarrollado y el mundo no desarrollado, entre el “norte global” y el “sur global”. La economía política de la inteligencia artificial (como parte del desarrollo económico vinculado a la tecnología en general) resulta un factor que no puede ser ignorado en el momento de diseñar soluciones normativas a los impactos negativos de su utilización.

Si bien el derecho internacional de los derechos humanos no es por sí solo –al menos en su dimensión normativa– suficiente para hacer frente a disparidades de desarrollo o a cuestiones profundas sobre justicia social, aún constituye, en nuestra opinión, una herramienta capaz de controlar la aplicación de desarrollos tecnológicos que sirvan al desarrollo humano, en la medida en que se releven tanto las dimensiones colectivas de los derechos humanos, así como el valor de los derechos económicos, sociales, culturales y ambientales en tanto parte de un sistema único e interconectado de derechos de igual valor, y las aristas políticas que permiten el pleno desarrollo de los derechos, como lo es (o puede ser) la democracia en sentido más amplio.



Una parte sustantiva del reclamo desde el mundo en desarrollo se refiere ya no a la capacidad de almacenamiento y procesamiento de la información, sino al control de la información misma, en particular de aquella información relativa a personas que denominamos “datos personales”, y a la capacidad del colectivo de ejercer efectivo control sobre esa información, más allá del control individual. En otras palabras, es clave para una efectiva gobernanza sobre la IA no solamente un marco de protección de datos personales (Veale et al., 2018), sino también la capacidad colectiva de reclamar control para el uso de esa información de manera agregada (Morozov, 2019), con la infraestructura institucional y política que habilite ese control. Además de esa arista normativa sobre los datos, otras formas de control sobre las propias empresas que ejercen su poder a través del diseño de sistemas de IA constituyen formas de acción estatal que pueden servir a los intereses individuales y colectivos dentro de un marco de pleno ejercicio de derechos humanos, idealmente no solo a nivel nacional o intraterritorial, sino también internacional.



3

Principios éticos para la inteligencia artificial

Como explican Taddeo y Floridi (2018), el desarrollo tecnológico es guiado y toma forma impulsado de manera variable por tres fuerzas normativas: la ética, la regulación, y la tecnología propiamente, con la ética cumpliendo funciones tanto en la formulación de regulación como en la interpretación de los enunciados normativos, y los tres elementos actuando de manera complementaria. Desde esta perspectiva, no es extraño que la ley por sí sola no sea el mecanismo de respuesta a desafíos de la tecnología; no obstante, no es del todo claro cómo funciona la interacción entre estas distintas fuerzas normativas, en función de distintas áreas de aplicación y de diferentes tecnologías.

A pesar de la comprensión de la multiplicidad de formas de abordar la gobernanza de la IA, es en el ámbito de la incorporación de la ética donde ha existido mayor desarrollo en los últimos años. Sin menospreciar su valor, corresponde examinarlo frente a una estructura de responsabilidad y de derechos establecida normativamente, y en contraste con esta última, su real o posible efectividad en el desarrollo y en la aplicación de IA.

3.1. Principios, marcos y directrices éticas

Examinando la discusión sobre la ética en los algoritmos, Mittelstadt et al. (2016) distinguían las preocupaciones epistémicas (relativas a la información procesada) y las preocupaciones normativas (relativas al resultado del proceso algorítmico) del uso de algoritmos, además del problema de la trazabilidad que cubre ambas dimensiones. El estudio reconoce a su vez la cadena entre datos que reflejan un sesgo, que llevan a decisiones que materializan discriminación, y los posibles efectos transformativos sobre la autonomía y la privacidad. Para Selbst y Barocas (2018), concentrarse en aspectos tales como la explicabilidad y la transparencia de las tecnologías puede desviar del valor de la intuición en el análisis de los aspectos normativos resultantes de una decisión algorítmica.

Un ejercicio de mapeo sistemático y de comparación de 36 conjuntos de principios y directrices vigentes hoy releva ocho temas comunes abordados por estos documentos: privacidad, rendición de cuentas (responsabilidad), protección y seguridad, transparencia y explicabilidad, equidad y no discriminación, control humano de la tecnología, responsabilidad profesional, y promoción de valores humanistas (Fjeld et al., 2020).

Un análisis distinto (Jobin et al., 2019) resume en cinco puntos la convergencia de la mayoría de los 84 documentos de principios y marcos éticos: transparencia, justicia y equidad, no maleficencia, responsabilidad, y privacidad. Sin embargo, existe divergencia sustantiva sobre la forma en que se interpretan los principios, su importancia, los temas, áreas o actores a que se refieren, y su forma de implementación.

Estas divergencias, si bien no son insalvables mediando suficiente acuerdo político entre las partes interesadas, revelan algunos vacíos significativos intrínsecos a los marcos éticos que pueden redundar en una mantención o en una amplificación de los riesgos de daño que buscan prevenir. En primer lugar, varios de los esfuerzos por incorporar aspectos éticos apuntan al desarrollo tecnológico en sus distintas etapas. Es el caso de aspectos tales como la explicabilidad o el manejo de datos. Curiosamente, quedan fuera otros aspectos éticos, tales como los usos intrínsecamente destructivos, como el caso de las armas automatizadas (Anderson y Waxman, 2013), o los efectos sobre la sociedad en la automatización de recomendaciones en redes sociales digitales, o las consecuencias medioambientales del desarrollo de la IA. La investigación respecto de las consecuencias sociales menos inmediatas se hace necesaria para modelar marcos éticos con una visión más completa de los riesgos que buscan prevenirse.

En segundo lugar, es también una omisión significativa en tales marcos éticos el abordaje de la falta de diversidad en la industria de avanzada de IA, que termina reforzando estereotipos culturales precisamente allí donde urge hacerse cargo de problemas de discriminación (Broussard, 2018; Crawford, 2016). En contraste con la abstracción de las empresas o instituciones que asumen compromisos éticos, la ausencia de responsabilidad o de consciencia en desarrolladores individuales sobre las consecuencias mediatas y distantes de los sistemas que desarrollan, alejan aún más una operatividad de esos marcos éticos.

En tercer lugar, en estrecha vinculación con lo anterior, la calificación de cierta cultura no es una preocupación meramente atingente a circunstancias económicas actuales. Es

expresión de un poder de explotación de la información personal a nivel masivo (Zuboff, 2019), que es a su vez extensión de un poder con raíces en el colonialismo y en el capitalismo (Coudry y Mejías, 2019). Aunque no es nuestro propósito actual discutir sobre los aspectos de la economía global de la IA, el análisis de esta relación a menudo conlleva puntos ciegos respecto de necesidades y de contextos de aplicación, de perspectivas de desarrollo económico, y en general, de las dimensiones colectivas en la aplicación de IA, en contraste con el énfasis en problemas de discriminación que afectan a individuos (aun agregados). Es más, como explicaremos más adelante, los marcos éticos, en atención a su origen, tienden a hacer propias las diferencias geográficas y políticas que ya ameritan el cuestionamiento de las tecnologías a las que se refieren.

3.2. Directrices éticas y la relación con los derechos humanos

Las directrices éticas y el derecho constituyen, para Gasser y Almeida (2017), niveles distintos de un modelo de tres capas para la gobernanza de la IA. Bajo este modelo, una primera capa basal es la referida a la dimensión técnica de los algoritmos y los datos (incluida su gobernanza), una segunda capa intermedia es la referida a la dimensión ética, y una tercera capa es legal y social, incluyendo normas y reglas, y construida esta sobre las anteriores. Este modelo teórico grafica una vinculación entre ética y derecho en el que este último construye pautas de conducta más específicas para la tecnología. Creemos, no obstante, que la relación no está construida necesariamente sobre niveles de abstracción de postulados normativos, sino más bien sobre la operatividad que el derecho provee, especialmente a través de la institucionalidad. En cuanto al contenido normativo sustantivo, por razones distintas de sus limitaciones de contenido, resulta necesario relevar el papel del derecho y, en particular, del derecho internacional de los derechos humanos, para proveer de un marco de gobernanza a la inteligencia artificial.



En primer término, ello es necesario en atención al origen de los marcos éticos. Para Calo (2017), las directrices éticas de la industria de la IA sugieren al Estado regulador que la autorregulación es suficiente para modelar la conducta de quienes desarrollan IA, sin ser necesario que existan leyes específicas para mitigar riesgos y eliminar escenarios de abuso. Como argumenta Hasselbalch (2019), las distintas iniciativas dan cuenta de espacios de negociación entre distintos grupos de interés que tratan de conducir la definición cultural de la ética de datos (y, por tanto, de sus usos), dentro de relaciones de poder complejas que se ejercen en el posicionamiento cultural. Parece una consecuencia lógica del hecho de que sean varias de las mismas empresas privadas de la industria de la IA las que han financiado iniciativas de ética en inteligencia artificial, postergando así un debate serio sobre la regulación legal de la IA (Nemitz, 2018). Esto posterga asimismo acciones estatales e internacionales concretas de búsqueda de responsabilidad frente a efectos nocivos sobre derechos humanos. Para Wagner (2019), esto corre el riesgo de constituir una forma falaz de asunción de compromisos éticos efectivos, evadiendo responsabilidades legales⁴.

En segundo lugar, es necesario relevar el rol del derecho frente a las directrices éticas en atención a los riesgos que ambos grupos de reglas buscan enfrentar. La discusión sobre la ética no ha ignorado los desafíos que presenta la inteligencia artificial para el derecho, incluidos los derechos humanos. No obstante, los marcos éticos descritos no logran abordar el espectro completo de

riesgos ni los desafíos de gobernanza, y tampoco cuentan con aceptación universal (Donahoe y Metzger, 2019), quizás como consecuencia de su proliferación en tanto productos de distintos actores interesados. Para Donahoe y Metzger (2019), es necesario un marco de gobernanza comprehensivo, sistemático y global, cuyo fundamento existe en el derecho internacional de los derechos humanos. La conveniencia se justifica en cuatro características del marco de derecho internacional de los derechos humanos: pone en el centro de la evaluación a la persona, cubre un amplio rango de preocupaciones, define roles y deberes de gobiernos y del sector privado, y tiene la ventaja de contar con apoyo global (Donahoe y Metzger, 2019). Si bien parece deseable contar con ese marco compartido de gobernanza normativa, su formulación volvería a buscar el desarrollo de directrices globales sin consideraciones regionales, como ocurre con las necesidades y perspectivas de desarrollo de países latinoamericanos. Aun fuera del problema de la gobernanza de la tecnología, algunos de los aspectos del derecho internacional de los derechos humanos ya son objeto de cuestionamiento por la falta de sensibilidad cultural. Estando a la espera de un consenso global contemporáneo por la urgencia por la incorporación de tecnologías de IA, volver los ojos a nuestra región resulta no solo deseable y sensible a nuestras diferencias, sino también práctico.

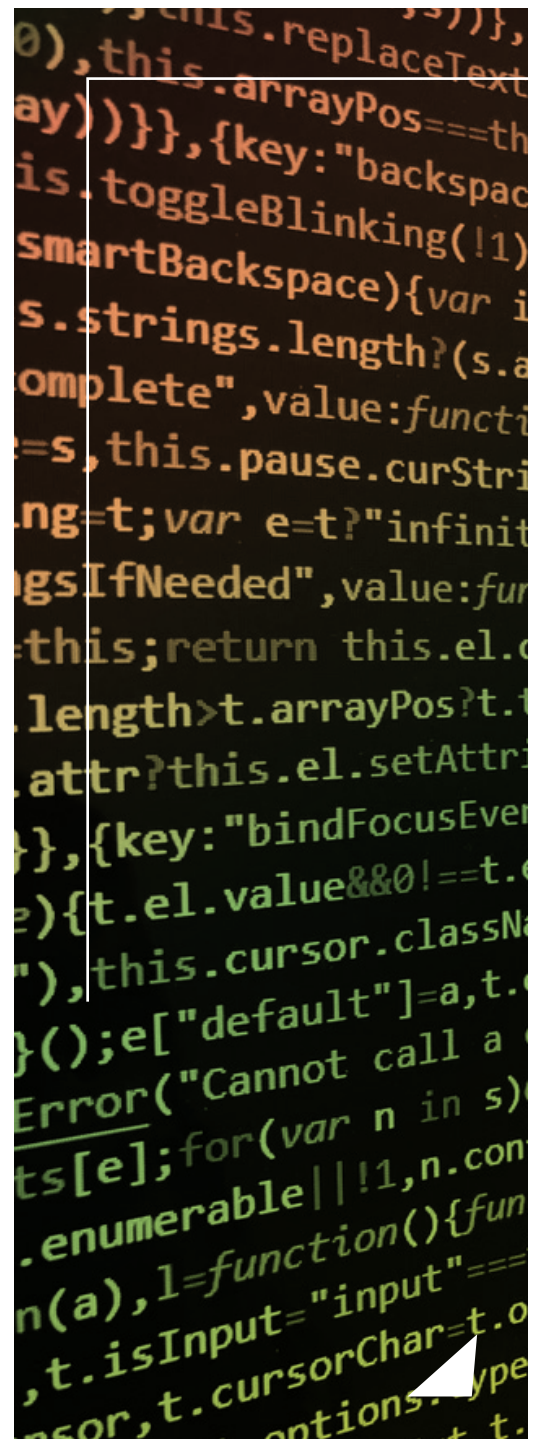
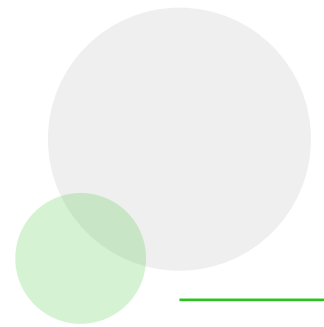
⁴ Por cierto, independientemente del origen de las iniciativas sobre directrices éticas para la IA, es necesario también reconocer que el inicio de una discusión dentro del contexto más reciente, desde una perspectiva regulatoria, existió también en el contexto europeo. Véase "Artificial Intelligence, Robotics, Privacy and Data Protection. Room document for the 38th International Conference of Data Protection and Privacy Commissioners", en el contexto de la ICDPPC de 2016 en Marrakech, disponible en: https://edps.europa.eu/sites/edp/files/publication/16-10-19_marrakesh_ai_paper_en.pdf

En tercer lugar, destaca la falta de mecanismos de observancia de las directrices éticas, que sean operados al menos en parte por personas distintas de quienes declaran hacer propios los principios o las directrices. Hasta el momento, se trata de compromisos voluntarios, carentes de instituciones y procedimientos de cumplimiento, y sin respaldo legal para hacerlos coercibles al menos frente a otros cocontratantes. Por un lado, el incumplimiento de estándares éticos no conlleva necesariamente efectos negativos o indeseados para quienes los incumplen, y ciertamente hace difícil la atribución de las infracciones éticas allí donde hay sistemas distribuidos de diseño y de desarrollo entre muchas personas. Por otro lado, el real impacto individual o colectivo derivado de una infracción de directrices éticas, de llegar a ocurrir, no encuentra en la propia ética mecanismos de remedio o de reclamo. Es decir, debido a que se centra en el cumplimiento de obligaciones preestablecidas, no asegura la rendición de cuentas por las consecuencias ulteriores, creando una brecha de responsabilidad (Liu y Zawieska, 2017).

En cuarto lugar, de manera práctica, no existe evidencia de la utilidad de marcos éticos en el desarrollo de tecnologías de inteligencia artificial. El compromiso de empresas privadas con principios éticos no significa necesariamente que toda su cadena productiva asuma obligaciones distintas en su operación cotidiana. Empíricamente, no hay demostración de que un marco ético afecte esa operación. McNamara et al. (2018), midiendo la influencia del código de ética para el desarrollo de software de la Association for Computing Machinery (ACM), un documento de 1972 actualizado en 2018, dieron cuenta de que no había evidencia de influencia significativa en la toma de decisiones de los desarrolladores de programas computacionales.

Esto último puede tener relación con la presencia de códigos éticos como un elemento añadido a una cadena que es capaz de operar sin él, una cadena a la que esos códigos son extraños, distintos de la producción misma en la industria de IA. En tanto marco de conducta sin reales consecuencias para personas individuales dentro de operaciones empresariales complejas, una serie de principios no parece suficiente para formar consciencia sobre las consecuencias sociales, ni sobre un sentido de responsabilidad. La ausencia de mecanismos de observancia ética, como existe en colegios profesionales, despoja a los principios éticos de real valor normativo.

En quinto lugar, como una extensión de lo recién expresado, la influencia efectiva de los marcos, principios o directrices éticas también debe ser relativizada en función de los grupos, instituciones, y miembros de la academia que se dedican a la ética en tecnologías de IA. Estos grupos están lejos de ser agentes de cambio con la capacidad de introducir formas de acción en las industrias de IA.



Por cierto, no se trata de aspectos que sean únicos de la ética en el uso de datos, algoritmos o inteligencia artificial, sino que son compartidos en la economía global con disímil participación en el desarrollo y en el despliegue de tecnología, con insuficientes niveles tanto de autorregulación como de regulación.

De este modo, si bien es positivo que existan intenciones y principios para la ética en algoritmos e IA, ello no puede servir para reemplazar o evitar regulaciones legales, coartando el proceso democrático (Nemitz, 2018).

Una arista enorme no cubierta por los principios éticos es la que responde a las desigualdades entre los países industrializados con participación fuerte en IA y los países en desarrollo, incluidos los de América Latina. Existen allí varias desigualdades relevantes, incluido el nivel de desarrollo económico y tecnológico. El impacto que la concentración del poder económico a nivel global tiene sobre el desarrollo puede incluso coartar el avance digital de países en desarrollo⁵. Como consecuencia, las desiguales condiciones de capital conllevan la reproducción de estructuras desiguales de oportunidad y de elección, y de la colectiva autodeterminación económica. Si la IA sirve para alcanzar el desarrollo –y, en consecuencia, si es deseable para los países en desarrollo–, la discusión sobre su gobernanza no puede limitarse al liberalismo y al individualismo que dominan el debate, cuyas ideas ignoran la dependencia que el capital hegemónico impone en individuos y colectivos mediante estructuras de elección reducidas en un más amplio orden global neoliberal (Gurumurthy y Chami, 2019). Abordar la explotación económica expandiendo elecciones individuales y colectivas, superando los problemas de sesgo y de discriminación, es un desafío que supera los actuales marcos éticos, y que agrega una dimensión compleja a una posible respuesta normativa.

4

Hacia la implementación normativa de una gobernanza ética

La suma y el traslado al terreno normativo y de políticas públicas de las consideraciones sobre ética y derechos humanos puede hacerse desde una perspectiva doble de la responsabilidad. Estas dimensiones son las correspondientes a la responsabilidad prospectiva, dirigida a prevenir y mitigar riesgos y amenazas, y a la responsabilidad histórica o retrospectiva, dirigida a asegurar con

⁵ Para la Conferencia de las Naciones Unidas sobre Comercio y Desarrollo, "la aparición de las prácticas monopolísticas por parte de las empresas principales y las plataformas digitales en todas las cadenas mundiales de valor, así como el ensanchamiento de la brecha digital, hace que pueda resultar extremadamente difícil para los países en desarrollo, especialmente los PMA [países menos adelantados], dar por sí solos el salto que les impulse hacia la industrialización digital" (UNCTAD, 2018).

justicia la atribución de responsabilidad ante el daño o la afectación de derechos fundamentales. Solamente si ambas dimensiones son recogidas, los individuos y la sociedad podrán confiar en que existen esfuerzos para prevenir daños, y mecanismos institucionales para asegurar la reparación y mitigar el daño ulterior (Yeung, 2019).

La gobernanza de sistemas algorítmicos en su arista de intervención estatal ofrece al menos cuatro opciones de política pública. Estas opciones son: la creación de consciencia mediante educación, monitoreo y denuncia; la responsabilidad en las decisiones algorítmicas en el sector público; la fiscalización regulatoria y la responsabilidad legal; y la coordinación global para la gobernanza (Koene et al., 2019). Si bien todas estas opciones reconocen un conjunto diverso de partes interesadas relevantes con distintos roles y responsabilidades, es crucial la participación de los estados, tanto en sus roles normativos a nivel nacional, como en sus roles de representación nacional en el contexto internacional, incluidos los foros intergubernamentales y los órganos de tratados de derechos humanos.

4.1. El rol del derecho

Poner el énfasis en desarrollar un marco de derechos humanos para la IA cumple al menos tres funciones. Primero, permite hacer frente a la falta de efectividad u observancia forzada de los marcos éticos. Segundo, permite enfrentar desde la capacidad de los estados latinoamericanos la falta de participación en la formulación de tecnologías de IA y de la formulación de los marcos éticos que a ella se refieren. Tercero, abre la posibilidad de un desarrollo propio de tecnologías con un anclaje más sólido en las consecuencias de derechos humanos. Esto último requiere hacerse cargo también de condiciones de desarrollo que parecen exceder el ámbito de acción de los derechos humanos; no obstante, existen diversas vías de cooperación que pueden desde ya ser adoptadas, al menos en lo relativo a IA, estrechamente relacionadas con los problemas relevados por la discusión sobre la ética (World Wide Web Foundation, 2017).

La regulación legal constituye asimismo una expresión normativa de la función estatal de protección de los intereses relevantes de sus ciudadanías. Como sostiene Nemitz (2018), una regulación legal reconoce que existe una dimensión normativa de los algoritmos, en tanto guías de la acción humana, pudiendo evaluarse esa dimensión de manera afín a la revisión legal a la luz de las normas superiores, tales como la Constitución y los instrumentos de derecho internacional de los derechos humanos. Adicionalmente, la regulación constituye una ventaja sustantiva frente a la ética y a la autorregulación, en atención a la concentración del poder

en un puñado de empresas: la legitimidad democrática y los mecanismos de observancia, estableciendo un terreno llano con orientación a la innovación e incentivos para su direccionamiento al interés público (Nemitz, 2018).

Esto no implica deshacerse de los marcos éticos. A pesar de las falencias aún no resueltas, la literatura sugiere con optimismo que los marcos éticos son útiles para hacer frente a riesgos propios de la IA. Incluso en América Latina puede relevarse ese rol: se atribuye acertadamente a la ética “un rol decisivo para el desarrollo de sistemas inclusivos y justos, poniendo límite a los sesgos que puedan encontrarse tanto en los datos como en la construcción de los modelos” (Mantegna, 2019). Las metodologías de prevención, aun si no pueden considerar claramente todo el espectro de oportunidades, riesgos y posibles consecuencias, pueden identificar alternativas preferibles y estrategias de mitigación de riesgos, bajando así el costo de oportunidad de no tomar decisiones u opciones por falta de claridad o por temor a una reacción negativa (Taddeo y Floridi, 2018). Sin embargo, es importante avanzar más allá de marcos éticos o de interpretaciones limitadas de equidad, responsabilidad y transparencia, abordando los aspectos de regulación y del modelo de negocios basado en datos, de manera holística (Cath, 2018).

Así, estos marcos éticos pueden representar una guía, en tanto directrices prácticas para la cadena de diseño, desarrollo e implementación de IA, especialmente en su uso dentro del aparato público. En clave de la prevención y del remedio de las afectaciones a los derechos humanos, se trata de mecanismos para las responsabilidades prospectiva y retrospectiva (Yeung, 2019). El contenido de esa transposición a un esquema normativo puede formularse en términos de obligaciones en distintos niveles y etapas, siendo un ejemplo concreto de propuesta favorable a esa transposición el requerimiento de conducción de evaluaciones de impacto sobre derechos humanos de forma previa a la implementación de determinadas tecnologías.

4.2. El reforzamiento de los derechos humanos a nivel internacional

Resulta inoficioso hacer un análisis de gobernanza de IA desde la perspectiva de los derechos humanos descuidando las relaciones de poder, en particular en un área en la que la concentración es tan evidente. La relación entre IA, democracia y derechos humanos requiere una visión holística sobre la realidad de la tecnología y los modelos de negocios, incluyendo la acumulación de poder tecnológico, económico y político en unas cuantas empresas, para el desarrollo de



IA en servicios comerciales (Nemitz, 2018). No se trata simplemente de la operación del marco de derechos humanos para afrontar riesgos adicionales derivados de esta estructura económica, sino de entenderla como un aspecto que requiere también un cambio en las propuestas de gobernanza de la IA que se hagan cargo de la profundización de las brechas de desarrollo y control, desde la mirada del constitucionalismo (Gurumurthy y Chami, 2019) y de los derechos fundamentales.

Como expone Yeung (2019), el poder y la sofisticación de las tecnologías digitales avanzadas en red pueden tomarse como base para entender que se trata de tecnologías radicalmente diferentes de otras, especialmente en razón de sus consecuencias en las estructuras sociales. Ello hace necesario fortalecer el discurso y los instrumentos de derechos humanos, relevando la responsabilidad colectiva de atender los fundamentos sociotécnicos de libertad moral y democrática, necesarios para la reclamación y la materialización de derechos humanos y de libertades fundamentales (Yeung, 2019). Los derechos humanos deben ser desarrollados para proteger a la humanidad contra los resultados de la robótica y la IA, fortaleciendo la noción de ser humano y el valor humano como algo intrínseco (Liu y Zawieska, 2017).

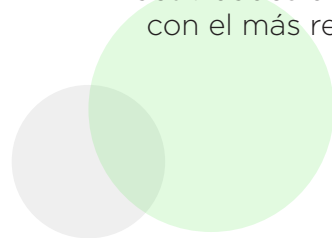
Parte de ese reforzamiento consiste en el respeto del conjunto de normas internacionales vigentes sobre derechos humanos por parte de actores privados, como las empresas dominantes a que hacemos referencia. Para Donahoe y Metzger (2019), un marco suficiente es provisto por la suma de la Declaración Universal de los Derechos Humanos y los Principios Rectores sobre las Empresas y los Derechos Humanos de Naciones Unidas, tanto para la gobernanza como para la evaluación de impacto de las decisiones de IA en las personas y la sociedad. Según los Principios Rectores, a la obligación estatal de proteger derechos humanos se suma la responsabilidad de respetar y proteger derechos humanos, y remediar sus afectaciones, en relación con sus productos, servicios y operaciones. Bajo los Principios Rectores,

las empresas deberían realizar evaluaciones de impacto en derechos humanos, consultar con partes interesadas y hacer públicos sus procesos de diligencia debida. Compartimos con Donahoe y Metzger (2019) la noción de la utilidad de estas normas y principios para el trabajo de deslindar responsabilidades entre gobiernos y empresas sobre los efectos de las tecnologías. No obstante, existen limitaciones intrínsecas a esta aproximación por la carencia de fuerza normativa para su observancia.

Por lo tanto, el reforzamiento implica también avanzar en el reconocimiento (esbozado anteriormente) de las consideraciones concretas sobre derechos fundamentales, sumando a un análisis instrumentos como la CADH, la COEA, el Pacto Internacional de Derechos Civiles y Políticos (PIDCP) y el Pacto Internacional de Derechos Económicos, Sociales y Culturales, apenas como punto de partida y sin ignorar otros instrumentos, especialmente aquellos que apuntan a la prohibición de discriminación y a la protección de individuos y grupos vulnerables. El carácter de la acción estatal en sistemas tecnológicos complejos merece ser objeto de materialización sustantiva y procedimental a nivel nacional, como mencionaremos más adelante. A la vez, los estados tienen como deuda el aseguramiento de mecanismos para la responsabilidad tanto prospectiva como retrospectiva, a nivel internacional, para lo cual los Principios Rectores son insuficientes. Otras iniciativas más recientes, como el marco ROAM-X de UNESCO⁶, y el informe antes referido del Consejo de Europa elaborado por Yeung (2019) apuntan en una dirección más concreta de acción estatal.

⁶ En el informe "Steering AI for knowledge societies", disponible en: <https://unesdoc.unesco.org/ark:/48223/pf0000372132>

Como hemos mencionado, esta necesidad de responsabilidad que cabe hacer valer a nivel internacional, es un imperativo naciente del poder acumulado (y creciente) de algunas de estas empresas, tanto por su insistencia en la acumulación de información, por la falta de transparencia de la tecnología que utilizan, y por la preterición sistemática de marcos legales nacionales, con la venia, resignación o indiferencia de gobiernos nacionales. Buena parte de esta preocupación puede ser abordada mediante un nuevo tratado internacional, como se ha discutido a partir de la creación por parte del Consejo de Derechos Humanos de Naciones Unidas de un Grupo de Trabajo Intergubernamental para la formulación de un “instrumento jurídicamente vinculante para regular, en el marco del derecho internacional de los derechos humanos, las actividades de las empresas transnacionales y otras empresas”⁷, con el más reciente borrador publicado en julio de 2019⁸.



4.3. Marcos normativos e implementación

A nivel latinoamericano, son varios los gobiernos de países de la región que han publicado documentación o al menos hecho anuncios en el sentido de guiar estratégicamente el desarrollo y la aplicación de la inteligencia artificial⁹. La experiencia de las agendas digitales muestra también que las prioridades nacionales tienden a cambiar significativamente con cada cambio de gobierno, reduciendo la probabilidad de mantener políticas de estado sostenidas en materia de tecnología y de coordinar la colaboración con vecinos en la región.

A ese respecto, y a pesar de la aparente falta de concreción o de evaluación de sus resultados, parece conveniente concertar una visión a nivel nacional o regional que integre el desarrollo tecnológico con el económico y el humano, según las propias aspiraciones de cada nación, y entendiendo sus particularidades y el contexto regional y global. Es decir, asumiendo el rol de la política en el terreno del desarrollo tecnológico, en vez de separarlo de ella. Para Cath et al. (2017), analizando comparativamente documentación sobre la visión del futuro de la IA en los Estados Unidos, la Unión Europea y el Reino Unido, hace falta una visión política global y una estrategia de largo plazo para el desarrollo de una “buena sociedad de IA” (“good AI society”). Cabe preguntarse si en los países de América Latina existen visiones de esa naturaleza, y qué legitimidad tendrían.

⁷ Consejo de Derechos Humanos, Resolución 26/9 sobre la Elaboración de un instrumento internacional jurídicamente vinculante sobre las empresas transnacionales y otras empresas con respecto a los derechos humanos (A/HRC/RES/26/9), 14 de julio de 2014. Disponible en: http://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/RES/26/9.

⁸ El borrador en inglés está disponible en: https://www.ohchr.org/Documents/HRBodies/HRCouncil/WGTransCorp/OEIGWG_RevisedDraft_LBI.pdf

⁹ Véase al respecto Aguerre (2020).

Independientemente del acuerdo político sobre una visión de la sociedad, la implementación local de reglas que den contenido normativo a la necesidad de responsabilidad prospectiva y retrospectiva por los sistemas de IA, resulta una tarea compleja. La regulación estatal, de existir, forma parte del ámbito de políticas públicas, en procesos que no es posible teorizar bajo un modelo único dadas las variaciones en contextos políticos y socioeconómicos diversos (Evans y Cvitanovic, 2018). No obstante, existen etapas no necesariamente lineales ni sucesivas, pero que surgen como oportunidades de intervención, como lo son: la formulación del problema, la formulación de una política, la toma de decisiones sobre instrumentos y recursos, la implementación de la política, y el monitoreo y evaluación de su funcionamiento (Evans y Cvitanovic, 2018). Tales instancias aparecen tanto respecto de políticas amplias sobre tecnología como de las medidas de implementación que ellas contienen.

La implementación en el campo normativo, por cierto, es una tarea desafiante. El Consejo de Europa, a mediados de 2019, conformó un Comité Ad-Hoc en Inteligencia Artificial (CAHAI, por sus siglas en inglés), cuyo propósito es examinar la factibilidad y los elementos potenciales de un marco legal para el desarrollo, diseño y aplicación de la inteligencia artificial, sobre la base de consultas amplias multisectoriales, basadas en los estándares del Consejo de Europa sobre derechos humanos, democracia y el estado de derecho¹⁰. El mandato del CAHAI se extiende hasta fines de 2021, con el deber de producir un estudio de factibilidad en los términos descritos. Aun cuando ese resultado es incierto, y se trata de un proceso iniciado dentro de la estructura y de la función propia del Consejo de Europa, es digno de atención como ejemplo del que pueden servirse otros países, incluidos los de América Latina, para la exploración de alternativas regulatorias.

Aspectos procedimentales y responsabilidad prospectiva

Siguiendo a Wagner (2018b), a propósito de requisitos para el diseño y desarrollo de la tecnología dentro de la ética o de los derechos humanos, como mínimo debe cumplirse con los siguientes criterios:

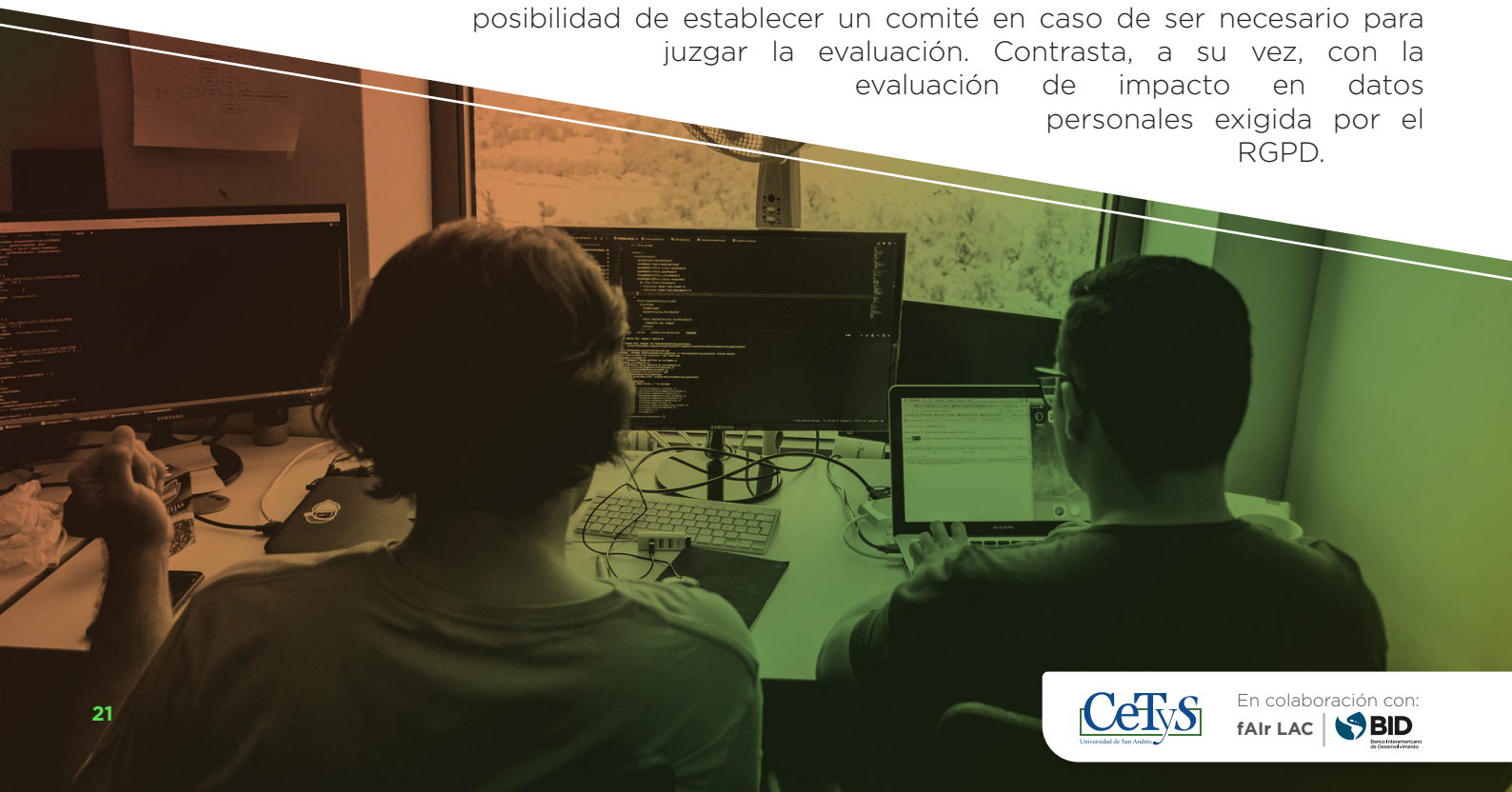
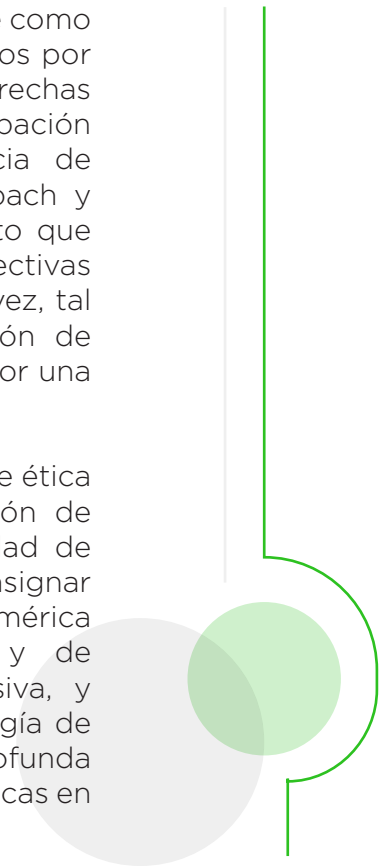
1. Participación externa: involucramiento temprano con partes interesadas relevantes;
2. Mecanismos de supervisión externa independiente (no necesariamente pública);
3. Procedimientos transparentes de justificación sobre por qué se adoptaron decisiones;
4. Lista de estándares no arbitrarios donde la selección de ciertos valores, principios éticos y derechos por sobre otros puede estar justificada plausiblemente;
5. Principios éticos que no sustituyan derechos humanos o fundamentales;
6. Declaración clara sobre la relación entre los compromisos y los marcos legales o regulatorios existentes, especialmente cuando estén en conflicto.

¹⁰ Términos de Referencia para el Comité Ad-Hoc sobre Inteligencia Artificial (CAHAI), aprobados por el Comité de Ministros del Consejo de Europa en decisión adoptada el 11 de septiembre de 2019 en la reunión de los Delegados de los Ministros N° 1353.

Tales aspectos parecen valiosos también en América Latina. Es clave como característica común en nuestra región que, a pesar de los esfuerzos por mejorar la calidad de la regulación en términos generales, existen brechas por superar, particularmente en relación con la falta de participación significativa de distintas partes interesadas, y con la ausencia de evaluaciones de impacto de la regulación, como apuntan Querbach y Arndt (2017). La necesidad de reglas para IA es quizás un ámbito que surge como una oportunidad para la aplicación de nuevas y efectivas formas de involucramiento y de participación multisectorial. A su vez, tal involucramiento requiere ser amplio: la deliberación, la formación de políticas públicas y la regulación de la IA deben estar informadas por una multiplicidad de disciplinas del conocimiento (Marda, 2018).

Latonero (2019) entrega una guía amplia para las consideraciones de ética y de derechos humanos en el desarrollo y en la implementación de inteligencia artificial, con especial consideración de la multiplicidad de actores interesados y de sus roles. No obstante, es necesario consignar que no todos los mismos actores son parte igualmente activa en América Latina, tradicionalmente excluida de la cadena de diseño y de programación de herramientas tecnológicas de aplicación masiva, y actualmente sin producción o exportación significativa de tecnología de IA. Por lo mismo, es también necesario examinar de manera más profunda la interdisciplinariedad como fuente de evidencia para políticas públicas en nuestros contextos.

Existen en la literatura varios esfuerzos para traducir a la práctica los marcos éticos existentes y mencionados. Un ejemplo es provisto por Mantelero (2018), que sugiere un modelo para un despliegue de inteligencia artificial respetuoso de derechos humanos centrado en los controladores o responsables de uso de datos. Mantelero propone el modelo de prevención de Evaluación de Impacto en Derechos Humanos, Ética y Sociedad (HRESIA, por sus siglas en inglés), para que los controladores evalúen valores en el diseño de productos y servicios, y la posibilidad de establecer un comité en caso de ser necesario para juzgar la evaluación. Contrasta, a su vez, con la evaluación de impacto en datos personales exigida por el RGPD.



El mecanismo de los análisis de impacto es uno de los que mejor se sirve del marco del derecho internacional de los derechos humanos. Atendido que los estados tienen el deber de prevención de impactos nocivos, su definición puede hacerse sobre la base del cuerpo de normas de derechos humanos, en un sentido amplio y comprensivo, y en relación con todas las etapas de la vida de los sistemas algorítmicos (McGregor et al., 2019). La incorporación de las características listadas por Wagner (2018b) y ya reseñadas puede enfocarse en ese rango de intereses.

Algunos de los aspectos de la responsabilidad prospectiva requieren detención adicional, en virtud de lo que significan para el desarrollo de sistemas de IA. Sin ánimo exhaustivo, por ser individualmente objeto de discusiones académicas y de búsqueda de soluciones técnicas, podemos destacar tres. Aun cuando la literatura tiende a examinarlos desde la perspectiva del sesgo y de la discriminación, el rango de impactos posibles es mayor.

El primero de los aspectos destacados es el de la transparencia de los sistemas algorítmicos, en el diseño, formulación, desarrollo, lógica interna de funcionamiento, e implementación, creando, según Pasquale (2015), “cajas negras” difíciles de escudriñar y de cuestionar. Esto puede referirse tanto a la complejidad técnica de su funcionamiento, como a la falta de revelación de su código amparada en reglas de propiedad intelectual (Noto La Diega, 2018), como a la difícil comprensión para personas sin formación en tecnología, ética o derecho. Sirve aquí como guía para inducir a una regulación efectiva sobre este punto el RGPD y el deber de informar, exigiendo que la entrega de información a los titulares de datos personales solicitantes sea entregada de forma “concisa, transparente, inteligible y de fácil acceso, con un lenguaje claro y sencillo”, permitiendo tomar acción sobre las decisiones o ejercer derechos de rectificación o cancelación¹¹.

La intervención humana es otro ámbito significativo, donde no basta con el cumplimiento de integrar más decisiones humanas para corregir sesgos o prevenir daños (“human in the loop”). Pasquale (2015) resalta la importancia de un sentido de agencia humana en procesos y sistemas de toma de decisiones, destacando múltiples ejemplos de impactos nocivos usualmente atribuidos a los algoritmos. Los argumentos para un rechazo a las decisiones plenamente automatizadas resultan bien resumidos por Noto La Diega (2018). Pero la noción de que integrar decisiones humanas en los procesos operados primordialmente por máquinas constituye resguardo suficiente de derechos humanos ha sido puesta bajo cuestionamiento, pudiendo incluso tener impacto negativo allí donde constituye más bien un gesto formal de humanización de las decisiones (Wagner 2018b). Si entendemos que existe un flujo continuo entre los sesgos de las personas, en razón de sus circunstancias materiales históricas, y los sesgos que incorporan los sistemas de decisión, una mayor interacción que detecte los efectos de esos sesgos no altera su existencia. De este modo, el nivel de agencia efectiva que tiene la intervención humana en decisiones automatizadas es un factor relevante para cumplir un rol de prevención de daños e, idealmente, para la promoción de un marco de derechos fundamentales en la implementación de un sistema de IA (Wagner, 2019).

¹¹ Para una discusión más completa, véase también Kaminski (2019).

Vinculado a lo anterior, otro aspecto relevante es el relativo a la participación en la formulación de las tecnologías, tanto respecto de la representatividad en las fuentes de información de las que se sirven los algoritmos de aprendizaje automatizado (es decir, los sesgos de sus bases de datos), como –más relevantemente– respecto de la existencia de procesos inclusivos y participativos en su desarrollo. En un sentido análogo al de los procesos de política pública (relevando el carácter normativo en la aplicación de IA), es necesaria la intervención de especialistas y de personas y comunidades posiblemente afectadas.

En definitiva, no es una aproximación satisfactoria la de convertir las guías y directrices en listados de cumplimiento. Además del cumplimiento formal estricto de normas y principios, es necesaria una incorporación de la ética de manera sensible a situaciones concretas, con directrices operativas para el proceso de toma de decisiones. Del mismo modo, reconociendo las limitaciones de una regulación externa, es necesario que la implementación técnica haga plena recepción de valores que parecen abstractos, aspecto de carácter formativo y organizativo que excede el propósito de este documento¹².

Actualización del marco normativo sustantivo relevante y responsabilidad retrospectiva

La necesidad de una regulación basada en derechos humanos para la gobernanza de la inteligencia artificial se dirige primordialmente al establecimiento de un marco de responsabilidad, fijando los mecanismos que debe atender esa regulación, en relación con los procesos que van desde la formulación hasta la implementación y evaluación. No obstante, es inoficioso concentrarse en el proceso sin consagración en lo sustantivo de derechos fundamentales, sus limitaciones, sus mecanismos de reclamación, sus formas de afectación (mediante actos definidos o que respondan a ciertas características), el facilitamiento de la búsqueda colectiva de responsabilidad, y el aseguramiento de la posibilidad de accionar contra empresas y contra entidades estatales.

En un sentido general, la calificación del daño, y consecuentemente del objeto de protección sujeto a riesgo, debe tomarse por los operadores del sistema de manera amplia. Mencionamos en el apartado 2.2. que existen esfuerzos en la literatura por listar derechos humanos potencialmente afectados; no obstante, su rango es más amplio que el representado por usuales preocupaciones por afectaciones individuales vinculadas a sesgos o a discriminación. Forman también parte de ellos, como hemos mencionado, consideraciones sobre democracia y autodeterminación, participación colectiva, y el medio ambiente. Una concepción más amplia del daño es provista por el derecho internacional de los derechos humanos, como señalan McGregor et al. (2019), al permitir una definición de daño concreta y universalmente aplicable, acompañada de tests sofisticados y desarrollados, incluyendo lo que constituye discriminación directa, indirecta e interseccional, además de sesgos estructurales e

¹² Sobre el particular, cabe mencionar la necesidad de integrar herramientas analíticas y técnicas para distintas comunidades potencialmente afectadas, no solamente en la educación, sino también para los trabajadores. Lo anterior incluye asimismo la capacidad de servir como contrapeso, lo que exige capacidad de organización y reconocimiento de sus intereses como grupos.



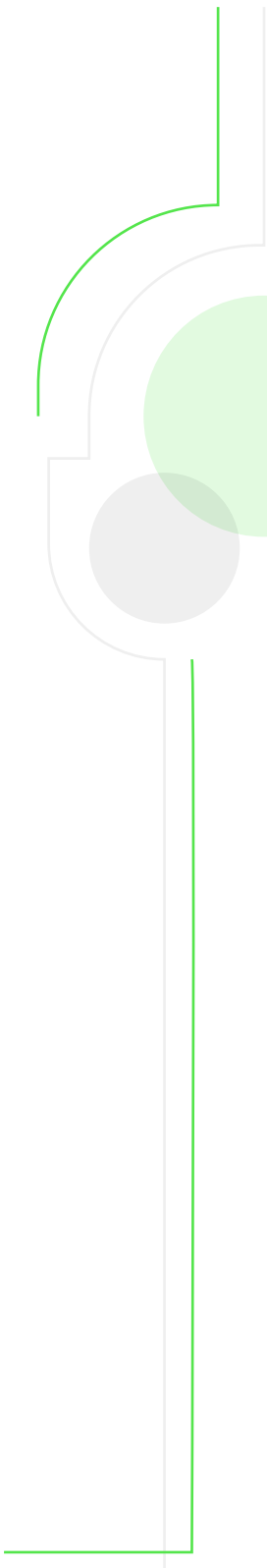
inconscientes, y extendiendo el análisis al impacto en los derechos de individuos y también de grupos. De ello se derivan consecuencias tales como la posible prohibición del uso de ciertos algoritmos, allí donde independientemente de cuestiones de sesgo puedan existir impactos significativos en derechos humanos (McGregor et al., 2019).

Como ya señalamos, los roles y las responsabilidades de estados y empresas, delineados por el derecho internacional de los derechos humanos, requieren implementación sustantiva y operativa a escala local. Puesto que los estados son los principales responsables de la protección efectiva de derechos humanos, tienen también la obligación de introducir marcos legislativos que reconozcan los derechos individuales y colectivos de las personas en sus territorios, así como las obligaciones y deberes legales de las entidades no estatales (Yeung, 2019).

Otro ámbito clave de regulación reside en la capacidad del ejercicio de los derechos vinculados a la protección de datos personales, especialmente allí donde se erige como una estructura normativa complementaria del ejercicio de derechos fundamentales. Esto exige, en parte, que tales marcos de protección existan, como ocurre ya con buena parte del mundo (Greenleaf, 2019), además de subsanar vacíos de cooperación y de coordinación (Aguerre, 2019). Requiere también la provisión de reformas que den contenido y operatividad a sus derechos en un contexto de algoritmos e IA, con el RGPD como un ejemplo habitual de regulación con aspectos sustantivos y operativos.

Sobre lo mismo, con particular pragmatismo, la Oficina del Comisionado de Información del Reino Unido entrega recomendaciones prácticas acerca de los análisis de “Big Data” que pueden extenderse en general a las tecnologías de procesamiento de información, en los sectores público y privado (Information Commissioner’s Office, 2017). Las distintas etapas de una evaluación de impacto en privacidad informativa surgen a la luz del cumplimiento de las normas vigentes a contar de la entrada en vigor del RGPD en 2018, cuando el Reino Unido era parte de la Unión Europea. La lección para América Latina sobre esto es doble: además de la necesidad de un marco fuerte de protección de información personal frente a la posibilidad de explotación o uso potencialmente dañino, la evaluación de los posibles riesgos sobre la privacidad informativa debe ser parte necesariamente embebida en las decisiones de implementación y de uso de herramientas avanzadas de procesamiento de datos. Es decir, tanto un marco acabado de protección (afín al RGPD), como una implementación guiada por autoridades públicas de control, útiles para estados y empresas. Es necesario reconocer, para el contexto latinoamericano, el reto que significa la pluralidad de reglas distintas en la materia, los desafíos de los mecanismos productivos de observancia y fiscalización, la promoción de programas de cumplimiento y responsabilidad, la cooperación entre autoridades nacionales y la interoperabilidad de los marcos legales, entre otros temas (Aguerre, 2019).

Además, la necesidad de regulación implica la toma de decisiones políticas sobre las consecuencias de responsabilidad penal y civil, que constituyen materias dignas de



exploración más profunda, separada de esta investigación. Sobre el particular, Yeung (2019) examina modelos de responsabilidad civil conforme a la clasificación entre responsabilidad estricta, responsabilidad condicionada e inmunidad. Scherer (2016) propone mecanismos de reducción de responsabilidad ulterior mediante la certificación del desarrollo de IA que se hagan cargo de medidas de responsabilidad prospectiva, en especial respecto de efectos no anticipados. En contra, McGregor et al. (2019) estiman que la imprevisibilidad, por ser conocida como tal por el actor que decide el uso de un algoritmo, no permite hacer declaraciones de responsabilidad reducida en abstracto, conforme al derecho internacional de los derechos humanos.

Finalmente, el establecimiento y el refinamiento de mecanismos de reclamación judicial resulta crucial. Además de constituir una forma de búsqueda de remedio contra empresas privadas, es un mecanismo para la reclamación de la propia responsabilidad estatal en la prevención del daño. El derecho internacional de los derechos humanos exige al estado medidas para prevenir la repetición del daño y para entregar reparación, además de hacer valer responsabilidades individuales. Por ello es clave que tanto estados como empresas provean mecanismos de reclamación, que atiendan a la asimetría de información entre las partes para facilitar el cuestionamiento de un proceso o resultado algorítmico (McGregor, 2019). Asimismo, la existencia de entidades o agencias, con procedimientos de reclamación asociados, puede servir como mecanismo eficiente y especializado (Scherer, 2016). En el mismo sentido, la litigación contra la acción estatal fundada en decisiones algorítmicas y de IA debe estar asegurada¹³. En general, el aspecto de la responsabilidad por los impactos de IA y algoritmos, en el sentido de la necesidad de rendición de cuentas, requiere todavía mayor exploración.

En síntesis, una actualización normativa adecuada a la era digital, a la situación del capitalismo global y al contexto de países en desarrollo y menos avanzados, puede servirse del derecho internacional de los derechos humanos como base sustantiva y procedimental. Además de definir las obligaciones estatales que cabe materializar internamente, entrega pautas para la definición y el examen de los daños y riesgos para individuos y comunidades que pueden crear los sistemas automatizados. Permite también la prevención mediante análisis de impacto basados en esas definiciones para todo el ciclo de formulación y utilización de IA, y construir esquemas de responsabilidad por la materialización de esos daños. Y de manera crucial, permite que ese análisis sirva a su vez para formular políticas que se dirijan a la materialización de los derechos individuales y colectivos, proveyendo a la tecnología de un rol reforzador de derechos fundamentales.

¹³ A modo de ejemplo, el AI Now Institute (2018) se centra en la litigación contra decisiones públicas delegadas a sistemas algorítmicos.

5

Conclusiones

El problema derivado de los potenciales riesgos de la toma de decisión algorítmica y del uso de inteligencia artificial, en especial en ámbitos de aplicación en el aparato público, ha sido discutido por gobiernos, instituciones internacionales, academia, comunidad técnica y sociedad civil, a través de la propuesta de distintos marcos de gobernanza y de la elaboración de principios y directrices éticas. Estas propuestas, que han proliferado en los últimos años, abordan variablemente la cadena de desarrollo y funcionamiento de esta familia de tecnologías, con diverso nivel de detalle y de especificidad.

Estos marcos éticos han sido criticados ampliamente por ser serviles a la actual situación de concentración de poder económico y tecnológico, tanto en razón de los orígenes de las propuestas como de la exclusión de consideraciones por parte de los intereses de personas de países menos desarrollados. Asimismo, tienen la debilidad de no contar con marcos de observancia práctica y rendición de cuentas, además de servir como fuente para evadir la discusión sobre regulación estatal e internacional efectiva. Aun si pudieren ser influyentes, incluso las instituciones y personas dedicadas al estudio de la ética en tecnología están lejos de ser agentes efectivos de cambio.

Sin embargo, los mismos marcos, en virtud de su multiplicidad y de los novedosos intentos por sistematizar su contenido y darles operatividad, sirven como fundamento inicial para una acción guiada por la sociedad o sus representantes en el Estado, para incorporar las consideraciones de derechos fundamentales con fuerza normativa a las distintas etapas de desarrollo de IA. Esto invita a una propuesta normativa de reforzamiento del derecho internacional de los derechos humanos, de la coordinación entre estados, y de la materialización de su contenido a nivel local.

Una propuesta de bases regulatorias no es en caso alguno una solución suficiente, aun si está enraizada en obligaciones vigentes bajo el derecho internacional de los derechos humanos. No es suficiente con abordar la prevención y el remedio de los riesgos, sin abocarse también a la economía política que es una fuente importante de esas amenazas y de la

creciente brecha de desarrollo entre países. No obstante, el derecho internacional de los derechos humanos entrega un marco para recoger al menos en parte esos otros aspectos, desde la perspectiva de los derechos de ejercicio colectivo, incluyendo los intereses por el desarrollo y por los derechos económicos, sociales, culturales y ambientales.

Una propuesta de bases regulatorias puede idealmente concentrarse en los mecanismos operativos de la gobernanza de la inteligencia artificial con respaldo en la ley y en la acción estatal. Es decir, en la incorporación fundamental de aspectos de derechos humanos para la regulación de la tecnología. Sin embargo, en países como los de América Latina, todavía es necesario asegurar la existencia de marcos normativos suficientemente protectores de la sustancia del amplio rango de derechos fundamentales, como ocurre con las reglas de responsabilidad civil y con la existencia de normas para la protección de datos personales, además de políticas tecnológicas sensibles a los contextos locales. La determinación de un marco de gobernanza de tecnologías que dé cuenta del interés fundamental por la autodeterminación y por el desarrollo, es un objetivo político que debe a su vez informar la acción regulatoria, relevando el interés público y los derechos de la comunidad.

A la vez, se hace necesario un esfuerzo internacional para establecer un contrapeso efectivo a las grandes empresas transnacionales de explotación de datos y tecnología. Independientemente de la existencia de marcos normativos internacionales en el área específica del procesamiento de datos, es necesario también fortalecer mecanismos de observancia forzosa en el ámbito internacional, como ocurre con la propuesta de un nuevo tratado sobre derechos humanos y empresas. Ello sostendría un margen de acción más robusto que el actualmente provisto únicamente por las normas del derecho internacional de los derechos humanos y los Principios Rectores sobre las Empresas y los Derechos Humanos de Naciones Unidas.

Reconociendo la diversidad de herramientas que el estado regulador puede utilizar para reforzar los aspectos potencialmente positivos de la IA, y prevenir sus riesgos, el anclaje en derechos humanos puede servir como guía. Aun si las herramientas pueden significar distintos niveles de regulación estatal o de autorregulación, de regulación nacional o internacional, de responsabilidades individualizadas o compartidas, se mantiene valor con el respaldo tanto del aparato estatal como de la comunidad internacional. Por cierto, ello exige condiciones de ejercicio democrático del poder para no constituir una forma adicional de despojo de control de las personas sobre sus vidas, a nivel individual y colectivo. En el mismo sentido, ello exige un ejercicio del poder estatal de una forma que no signifique, como ocurre en un sinnúmero de ejemplos, una actitud represiva frente a la tecnología, ni una instrumentalización del poder para, a través de la regulación de la tecnología, asegurar y proteger los objetivos políticos de la autoridad.

El derecho internacional de los derechos humanos, sirviéndose de su propio desarrollo por más de medio siglo, sumado a las consideraciones más recientes vinculadas a la tecnología, provee un enfoque basado en

derechos fundamentales para abordar la formulación, el diseño, el desarrollo, el despliegue, la aplicación y la evaluación de decisiones y de sistemas de IA. Sirve de guía para identificar patrones de riesgo y responsabilidades entre agentes estatales y no estatales, incluidas las empresas transnacionales. Y no excluye la existencia de códigos y directrices con base en la ética para un desarrollo positivo de las herramientas tecnológicas, especialmente allí donde la regulación pudiere ser insuficientemente detallada y a modo de interpretación.

Puesto que el avance tecnológico, su influencia en el aparato público, y sus consecuencias en la vida de las personas y de las comunidades son particularmente vertiginosas, la tarea de proveer marcos regulatorios para nuevas tecnologías que fomenten usos positivos y prevengan riesgos resulta una tarea urgente.

Referencias bibliográficas

Agrawal, A., Gans, J. y Goldfarb, A. *Prediction Machines: The simple economics of artificial intelligence*, Harvard Business Review Press, Boston, 2018.

Aguerre, C. Digital trade in Latin America: mapping issues and approaches, en *Digital Policy, Regulation and Governance*, vol. 21 n. 1, 2019, pp. 2-18. <https://doi.org/10.1108/dprg-11-2018-0063>

Aguerre, C. *Estrategias nacionales de IA y gobernanza de datos* (en prensa). Universidad de San Andrés, Buenos Aires, 2020.

AI Now Institute. *Litigating algorithms: Challenging government use of algorithmic decision systems*, 2018. Disponible en: <https://ainowinstitute.org/litigatingalgorithms-2019-us.pdf>

Anderson, K. y Waxman, M. C. *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*, Stanford University, The Hoover Institution (Jean Perkins Task Force on National Security and Law Essay Series); American University, WCL Research Paper 2013-11; Columbia Public Law Research Paper, 2013, pp. 13-351. <http://dx.doi.org/10.2139/ssrn.2250126>

Angwin, J., Larson J., Mattu, S. y Kirchner, L. *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks*, ProPublica, 2016. Disponible en: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

Artificial Intelligence for Development. *Government Artificial Intelligence Readiness Index*, 2019. Disponible en: <https://ai4d.ai/index2019>.

Broussard, M. *Artificial Unintelligence: How Computers Misunderstand the World*. MIT Press, Cambridge, 2018.

Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., Henke, N. y Trench, M. *Artificial Intelligence: the Next Digital Frontier?*, McKinsey Global Institute, 2017.

Calo, R. Artificial Intelligence Policy: A Primer and Roadmap, en *Davis Law Review*, University of California, vol. 51, 2017, pp. 399-435.

Campolo, A., Sanfilippo, M., Whittaker, M., y Crawford, K. *AI Now 2017 Report*. AI Now Institute, 2017. Disponible en: https://ainowinstitute.org/AI_Now_2017_Report.pdf

Cath, C. Governing artificial intelligence: ethical, legal and technical opportunities and challenges, en *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, n. 376(2133), 2018. <http://doi.org/10.1098/rsta.2018.0080>

Cath, C., Wachter, S., Mittelstadt, B. D., Taddeo, M. y Floridi, L. Artificial Intelligence and the "Good Society": the US, EU, and UK approach, en *Science and Engineering Ethics*, 538(7625), 2017, pp. 1-24. <http://doi.org/10.1007/s11948-017-9901-7>

Citron, D. K. y Jurecic, Q. *Platform Justice: content moderation at an inflection point*, Hoover Institution, Aegis Series Paper No. 1811, 2018.

Couldry, N. y Mejías, U. Data colonialism: rethinking big data's relation to the contemporary subject, en *Television and New Media*, vol. 20 n. 4, 2019, pp. 336-349.

Crawford, K. Artificial Intelligence's White Guy Problem, en *The New York Times*, 26 de junio de 2016. Disponible en: <https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>.

Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., Kak, A., Mathur, V., McElroy, E., Nill Sánchez, A., Raji, D., Rankin, J. L., Richardson, R., Schultz, J., Myers West, S., Whittaker, M. (2019). *AI Now 2019 Report*. Nueva York: AI Now Institute. Disponible en: https://ainowinstitute.org/AI_Now_2019_Report.html.

Daly, A., Hagendorff, T., Hui, L., Mann, M., Marda, V., Wagner, B., Wang, W. y Witteborn, S. *Artificial Intelligence Governance and Ethics: Global Perspectives* (Research Paper No. 2019-15), Chinese University of Hong Kong, Hong Kong, 2019.

Donahoe, E. y Metzger, M. M. Artificial Intelligence and Human rights, en *Journal of Democracy*, vol. 30 n. 2, 2019, pp. 115-126.

Elish, M. C. y Boyd, D. Situating methods in the magic of Big Data and AI, en *Communication Monographs* vol. 85 n. 1, 2017, pp. 57-80. <http://doi.org/10.1080/03637751.2017.1375130>

Eubanks, V. *Automating inequality: how high-tech tools profile, police, and punish the poor*, St. Martin's Press, Nueva York, 2018.

Evans, M.C. y Cvitanovic, C. An introduction to achieving policy impact for early career researchers, en *Palgrave Commun* vol. 4, 88, 2018. <http://doi.org/10.1057/s41599-018-0144-2>

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. C. y Srikumar, M. *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI, 2020* [online]. Disponible en: <https://cyber.harvard.edu/publication/2020/principled-ai>

Gasser, U. y Almeida, V. A Layered Model for AI Governance, en *IEEE Internet Computing* vol. 21 n. 6, 2017, pp. 58-62. <http://doi.org/10.1109/mic.2017.4180835>

Greenleaf, G. Global Data Privacy Laws 2019: 132 National Laws & Many Bills, en *Privacy Laws & Business International Report n. 157*, 2019, pp. 14-18. Disponible en: <https://ssrn.com/abstract=3381593>

Gurumurthy, A. y Chami, N. The Wicked Problem of AI Governance, en *Artificial Intelligence in India, vol. 2*, Friedrich Ebert Stiftung, 2019. Disponible en: <http://library.fes.de/pdf-files/bueros/indien/15763.pdf>

Hasselbalch, G. Making sense of data ethics. The powers behind the data ethics debate in European policymaking, en *Internet Policy Review*, vol. 8 n. 2, 2019. <http://doi.org/10.14763/2019.2.1401>

Information Commissioner's Office. *Big data, artificial intelligence, machine learning and data protection*, 2017.

Jobin, A., Ienca, M. y Vayena, E. The global landscape of AI ethics guidelines, en *Nature Machine Intelligence*, vol. 1 n. 9, 2019, pp. 389-399. <http://doi.org/10.1038/s42256-019-0088-2>

Kaminski, M. E. The Right to Explanation, Explained, en *Berkeley Technology Law Journal*, vol. 34, 2019, pp. 189-218. <http://doi.org/10.15779/Z38TD9N83H>

Koene, A., Clifton, C., Hatada, Y., Webb, H., Patel, M., Machado, C., LaViolette, J. y Richardson, R. *A governance framework for algorithmic accountability and transparency*, European Parliamentary Research Service, 2019.

Latonero, M. *Governing Artificial Intelligence: Upholding Human Rights and Human Dignity*. *Data & Society*, 2019. Disponible en: https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf.

Liu, H.-Y. y Zawieska, K. From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence, en *Ethics and Information Technology*, vol. 19 n. 3, 2017. <http://doi.org/10.1007/s10676-017-9443-3>

Mantegna, M. OK Computer: ¿de qué hablamos cuando hablamos de IA?, en: I. Petrella, P. Marzocca y A. Beun (compiladores), *Ideas para la Argentina del 2030*, Área Cuatro, Buenos Aires, 2019.

Mantelero, A. AI and Big Data: A blueprint for a human rights, social and ethical impact assessment, en *Computer Law & Security Review*, vol. 34 n. 4, 2018, pp. 754-772. <http://doi.org/10.1016/j.clsr.2018.05.017>

Marda, V. Artificial intelligence policy in India: a framework for engaging the limits of data-driven decision-making, en *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 2018. <http://doi.org/10.1098/rsta.2018.0087>

McGregor, L., Murray, D. y Ng, V. International Human Rights Law as a framework for algorithmic accountability, en *International and Comparative Law Quarterly*, vol. 68 n. 2, 2019, pp. 309-343. Disponible en: <http://doi.org/10.1017/S0020589319000046>

McNamara, A., Smith, J. y Murphy-Hill, E. *Does ACM's code of ethics change ethical decision making in software development?*, Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE '18), 2018, pp. 729-733. <http://doi.org/10.1145/3236024.3264833>

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. y Floridi, L. The ethics of algorithms: Mapping the debate, en *Big Data & Society*, vol. 3 n. 2, 2016, pp. 1-21. <http://doi.org/10.1177/2053951716679679>

Morozov, E. Digital Socialism? The calculation debate in the age of big data, en *New Left Review*, n. 116/117, 2019, pp. 33-67.

Muñoz León, F. Epistemología de la techne: a propósito del fraude informático, en *Revista Chilena de Derecho y Tecnología*, vol. 2 n. 2, 2013, pp. 247-260. <http://doi.org/10.5354/0719-2584.2013.30314>. Disponible en: <https://rchdt.uchile.cl/index.php/RCHDT/article/view/30314>

Nemitz, P. Constitutional democracy and technology in the age of artificial intelligence, en *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376, 2018. <http://doi.org/10.1098/rsta.2018.0089>

Noble, S.U. *Algorithms of oppression: how search engines reinforce racism*. New York University Press, Nueva York, 2018.

Noto La Diega, G. Against the Dehumanisation of Decision-Making – Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information, en *JIPITEC*, vol. 9 n. 1, 2018, pp. 3-34.

O'Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown Publishers, Nueva York, 2016.

Pasquale, F. *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge, MA, 2015.

Querbach, T. y Arndt, C. *Regulatory policy in Latin America: An analysis of the state of play*. OECD Regulatory Policy Working Papers, No. 7, OECD Publishing, París, 2017. <http://dx.doi.org/10.1787/2cb29d8c-en>

Raso, F., Hilligoss, H., Krishnamurthy, V., Bavitz, C. y Kim, L. *Artificial Intelligence & Human Rights: Opportunities and risks*, Berkman-Klein Center for Internet & Society, Harvard University, 2018.

Scherer, M. Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies, en *Harvard Journal of Law & Technology*, vol. 29 n. 2, 2016, pp. 353-400.

Selbst, A. D. y Barocas, S. The intuitive appeal of explainable machines, en *Fordham Law Review*, vol. 87, 2018, pp. 1085-1139.

Taddeo, M. y Floridi, L. How AI Can Be a Force for Good, en *Science*, n. 361 (6404), 2018, pp. 751-52.

UNCTAD. *Informe sobre el comercio y el desarrollo*, 2018 [online]. Disponible en: https://unctad.org/es/PublicationsLibrary/tdr2018_es.pdf.

Veale, M., Binns, R. y Edwards, L. Algorithms that remember: model inversion attacks and data protection law, en *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376 (2133), 2018. <http://doi.org/10.1098/rsta.2018.0083>

Wagner, B. *Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications*, Consejo de Europa, Comité de Expertos en Intermediarios de Internet (MSI-NET), 2018a.

Wagner, B. Ethics as an Escape From Regulation: From 'Ethics-Washing' to Ethics-Shopping?, en *Bayamlioğlu, E., Baraliuc, I., Janssens, L.A.W. y Hildebrandt, M.* (editores), *Being Profiled: Cogitas Ergo Sum*, Amsterdam University Press, Amsterdam, 2018b, pp. 84-89).
<https://doi.org/10.2307/j.ctvhrd092.18>

Wagner, B. Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems, en *Policy & Internet*, vol. 11 n. 1, 2019, pp. 104-122. <http://doi.org/10.1002/poi3.198>

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S.M., Richardson, R., Schultz, J. y Schwartz, O. *AI Now Report 2018*. AI Now Institute, 2018. Disponible en:
https://ainowinstitute.org/AI_Now_2018_Report.pdf.

World Wide Web Foundation. *Artificial Intelligence: The Road Ahead in Low and Middle-Income Countries*, 2017.

Yeung, K. *A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework*, MSI-AUT(2018)05 rev. Consejo de Europa, Comité de Expertos en dimensiones de derechos humanos del procesamiento automatizado de datos y distintas formas de inteligencia artificial (MSI-AUT), 2019 [online]. Disponible en:
<https://rm.coe.int/a-study-of-the-implications-of-advanced-digital-technologies-including/168094ad40>.

Zuboff, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, PublicAffairs, Nueva York, 2019.

Descargo de responsabilidad. Las opiniones expresadas en la publicación incumben únicamente a los/as autores/as. No tienen intención de reflejar las opiniones o perspectivas del CETyS ni de ninguna otra organización involucrada en el proyecto.