

# La gobernanza de la Inteligencia Artificial en América Latina: entre la regulación estatal, la privacidad y la ética digital

Castaño, D. (2020). La gobernanza de la Inteligencia Artificial en América Latina: entre la regulación estatal, la privacidad y la ética digital. En C. Aguerre, (Ed.). *Inteligencia Artificial en América Latina y el Caribe. Ética, Gobernanza y Políticas*. Buenos Aires: CETyS Universidad de San Andrés.

Daniel Castaño\*

## Resumen

*El presente escrito tiene por objeto analizar diferentes modelos de gobernanza de la inteligencia artificial en América Latina. Puntualmente, se enfocará en el derecho a obtener explicaciones sobre decisiones individuales automatizadas (“right to explanation”) toda vez que este derecho se encuentra en la intersección de los derechos a la privacidad y autonomía personal. Para tal efecto, primero revisará, en términos generales, los fundamentos normativos y teóricos del derecho a obtener explicaciones en Europa y, con base en ello, analizará si dicho derecho existe dentro del contexto latinoamericano y la manera en qué la ética digital jugará un papel determinante en la definición de sus contornos normativos e implementación práctica.*

\* Profesor de Derecho de la Universidad Externado de Colombia y abogado de la misma institución. LL.M y J.S.D de las University of California, Berkeley. Consultor independiente en temas de derecho, tecnología y ética digital.

## Introducción

En una conversación reciente con Fei-Fei (<https://profiles.stanford.edu/fei-fei-li>) Li en la Universidad de Stanford, Yuval Noah Harari (<https://www.ynharari.com/>) se apoyó en la siguiente ecuación para explicar la forma en que ciertos conceptos filosóficos tradicionales están siendo cuestionados por las tecnologías disruptivas: **conocimiento biológico x poder de computación x datos = habilidad de “hackear” a los humanos**. Para evitar que ello suceda, Harari propone que todas las personas tengamos un sistema personal de inteligencia artificial que nos sirva a nosotros y a nuestros intereses personales por encima de los intereses corporativos o gubernamentales. Si bien dicha ecuación y las propuestas de Harari carecen por ahora de todo fundamento científico, sus planteamientos suscitan complejas preguntas jurídicas y éticas que se deben resolver en el corto plazo para asegurar que la inteligencia artificial se encuentre al servicio de la humanidad y no al contrario.

Por su parte, Stephen Hawking postuló que la creación de la inteligencia artificial podría ser el suceso más importante en la historia de la humanidad (Hawking, 2018). En este sentido, la literatura sostiene que la inteligencia artificial tiene el potencial de transformar radicalmente las relaciones sociales, culturales, económicas, políticas y legales (Domingos, 2015). En efecto, sus avances podrían optimizar o aumentar la inteligencia humana, tanto en el ámbito individual como colectivo, para promover la creatividad, diversificar el pensamiento y robustecer la eficiencia en la realización de muchas tareas que ocupan nuestra cotidianidad personal y laboral (Goldberg y Kumar, 2018; Jordan, 2019). Pero, como bien lo advierten Barrat (2013) y Bostrom (2014), la inteligencia artificial también podría ser nuestro último invento si no aprendemos a evitar y a controlar los riesgos que se derivan de su rápido desarrollo e implementación.


De ahí que en el presente escrito no se aborden las preocupaciones que genera para algunos el eventual desarrollo de lo que Bostrom (2014) denomina “superinteligencia” (también conocida como IA general o fuerte), entendida como un sistema de inteligencia artificial general que podría exceder todas las capacidades cognitivas humanas en virtualmente cualquier disciplina. Se trata más bien de concentrarse en los retos jurídicos y éticos que conlleva la inteligencia artificial débil o acotada. Esto especialmente en lo que se refiere a los derechos humanos, la democracia y el Estado de Derecho, y a los efectos que aquella pueda tener en la privacidad, agencia, autonomía y autodeterminación de los individuos (Katyal, 2019 y Barlett, 2019).

El evento académico se puede consultar en <https://www.youtube.com/watch?v=d4rBh6DBHyw>

No es esta la primera vez que la humanidad se enfrenta al reto de regular un asunto complejo dentro de un ambiente de incertidumbre. Esta situación se ha repetido en cada una de las revoluciones industriales, dando lugar a transformaciones regulatorias en distintos ámbitos. En el presente documento se busca exponer los principales retos jurídicos y éticos de la inteligencia artificial, para luego, con base ello, sugerir cuál sería la estrategia regulatoria más adecuada para gestionar dichos retos en América Latina, en particular aquellos asociados con la privacidad y autonomía de los individuos.

De acuerdo con la doctrina, la privacidad, agencia, autonomía y autodeterminación son derechos que se encuentran en una intersección donde confluyen principios éticos y normas jurídicas con las cuales se busca intervenir en un ámbito de privacidad y autonomía personal para lograr dos fines: (i) proteger la información personal en la forma, grado y extensión en que el individuo decide compartirla con otros, y (ii) proteger igualmente la forma libre y autónoma en que este toma sus decisiones personales con base en dicha información sin ningún tipo de injerencia externa no consentida, indebida o ilegal.

Dada la extensión del tema, así como las limitaciones de espacio del presente documento, aquí se hará un énfasis particular en el derecho a obtener explicaciones sobre decisiones individuales automatizadas (*right to explanation*). Ello partiendo del postulado de que este derecho se encuentra en la intersección del derecho a la privacidad y autonomía personal, y que apunta a garantizar que el individuo conozca la manera en que sus datos personales son procesados de manera automatizada por sistemas de IA, a partir de lo cual podrá tomar decisiones que se ajusten a su plan de vida sin ningún tipo de injerencia externa ilegal, indebida o no consentida.



Para tal efecto, primero se explicará el concepto de inteligencia artificial, su fuerza disruptiva y las diferentes definiciones que han sido acogidas por la literatura y por diferentes instrumentos internacionales. Luego se presentarán los principales problemas de la gobernanza de los sistemas de inteligencia artificial relacionados con la transparencia, responsabilidad, control y explicabilidad algorítmica.

Con base en dichas consideraciones se describirán los principales problemas jurídicos y éticos relacionados con el procesamiento automatizado de datos personales mediante sistemas de inteligencia artificial. Posteriormente se reseñarán, en términos generales, los fundamentos normativos y teóricos del derecho a obtener explicaciones en Europa y, con base en ello, se analizará si dicho derecho existe dentro del contexto latinoamericano y la manera en que la ética digital desempeñará un papel determinante en la definición de sus contornos normativos e implementación práctica.

Por último, se abordará la importancia de la implementación de iniciativas de experimentación regulatoria (*regulatory sandboxes*) en América Latina para definir la naturaleza, extensión, contornos y aplicación práctica del derecho a obtener explicaciones a partir de la experimentación tecnológica e innovación regulatoria por medio de instrumentos regulatorios alternativos (véase Castaño, 2019).



## 2

## La Inteligencia Artificial y su fuerza disruptiva

Como se indicó en la sección anterior, el presente estudio se enfocará en los retos éticos y jurídicos derivados del diseño, desarrollo e implementación de sistemas de inteligencia artificial débil o acotada, por oposición a los sistemas generales o fuertes o AGI (véase Russell, 2017). Para ello es pertinente tener como referencia las definiciones de “sistemas de inteligencia artificial” adoptadas por el Grupo de Expertos de Alto Nivel (HLEG por sus siglas en inglés) de la Unión Europea sobre Inteligencia Artificial y por la OCDE, las cuales han sido ampliamente aceptadas en el ámbito internacional. En el informe publicado el 8 de abril de 2019, este grupo de expertos definió los sistemas de inteligencia artificial así:

“Los sistemas de IA son software (y posiblemente hardware también) diseñados por humanos que, de acuerdo con un objetivo complejo determinado, actúan dentro de una dimensión física o digital percibiendo su ambiente por medio de la recolección de datos, interpretando datos estructurados o no estructurados, razonando sobre el conocimiento o procesando la información derivada de dichos datos para tomar las mejores acciones que le permitan cumplir el objetivo que le ha sido asignado” (HLEG, 2019).

Por su parte la OCDE, en las Recomendaciones sobre Inteligencia Artificial el 22 de mayo de 2019, en su numeral I, define los sistemas de inteligencia artificial en los siguientes términos:

“Un sistema de IA es una máquina que puede, de acuerdo con un conjunto de objetivos definidos por humanos, realizar predicciones, recomendaciones o tomar decisiones que tengan una influencia sobre ambientes reales o virtuales. Los sistemas de IA son diseñados para operar con distintos niveles de autonomía” (OCDE, 2019a).

En este contexto, la fuerza disruptiva de la inteligencia artificial se encuentra condicionada por tres variables esenciales: el aumento exponencial en el poder computacional; la sofisticación y proliferación de

algoritmos de código abierto; y por último, aunque tal vez lo más importante, la generación diaria de billones de gigabytes. No en vano los datos han sido calificados como la materia prima de la Cuarta Revolución Industrial, hasta el punto de que algunos estiman que actualmente su valor es más alto que el del petróleo (Stephens-Davidowitz, 2017). En este sentido, la OCDE ha señalado que la recolección y procesamiento de una miríada de datos dio lugar al surgimiento y fortalecimiento del ecosistema conocido como inteligencia de datos, macrodatos o *Big Data* (OECD, 2015).

De una parte, los datos pueden ser de naturaleza personal, como por ejemplo aquellos asociados a una persona identificada o identificable a partir de sus hábitos, comportamientos o movimientos. De otra parte, tal información puede ser de naturaleza institucional, como por ejemplo la relativa a la salud pública, la propiedad y los tributos (Mittelstadt y Floridi, 2016). Según la OCDE, los usuarios de diferentes plataformas y servicios digitales proporcionan inteligencia de datos a las corporaciones para su procesamiento, lo que les permite automatizar sus procesos, experimentar, y crear nuevos productos y modelos de negocio (OECD, 2015). En ese sentido, el Departamento del Tesoro del Reino Unido ha señalado que los datos permiten el desarrollo de nuevos modelos de negocios digitales que moneticen las reacciones del usuario (*user engagement*) y la transformación de los servicios a cargo del Estado (HM Treasury, 2018).

Es por ello que tanto los actores públicos (OECD, 2019b) como los privados (McKinsey Global Institute, 2017 y Shaw, 2019) sostienen que la inteligencia artificial y la inteligencia de datos no solo están impulsando la transformación de muchos sectores productivos sino también la disrupción digital de varias industrias y de la economía digital, lo cual genera billones en ganancias. Además, la inteligencia artificial también podría tener un alto valor e impacto social, en la medida en que podría ayudar a combatir la desigualdad, la corrupción, el crimen y el cambio climático, promover la justicia social y mejorar la calidad de vida de la población (Floridi et al., 2018 y OECD, 2019b). En otras palabras, los datos tienen un importante valor social y económico cuyo procesamiento por medio de sistemas de inteligencia artificial tiene el potencial de impulsar radicalmente la economía digital, crear nuevos modelos de negocio, y transformar al Estado junto con los servicios a su cargo.





## La gobernanza de la inteligencia artificial: entre la regulación estatal, la privacidad y la ética digital

El procesamiento de datos personales e institucionales mediante sistemas de inteligencia artificial para diagnosticar, describir, predecir o prescribir información crea problemas jurídicos y éticos en tres niveles distintos, aunque estrechamente ligados con los desafíos que impone la gobernanza de los sistemas de inteligencia artificial: (i) transparencia y responsabilidad algorítmica, (ii) control de los sistemas de IA, y (iii) explicabilidad e inteligibilidad algorítmica,

Con base en la abundante literatura en la materia, es posible afirmar que los principales retos jurídicos y éticos de los sistemas de

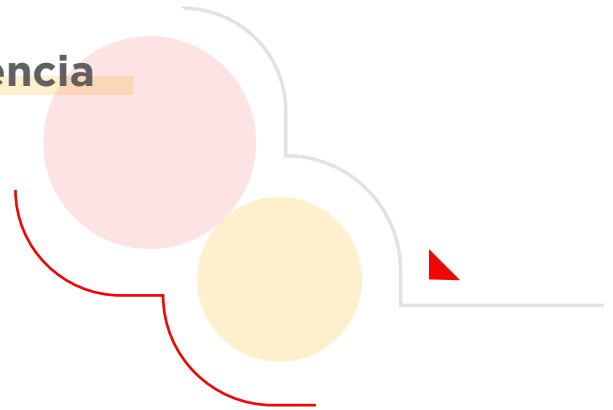
inteligencia artificial derivan, en esencia, de situaciones relacionadas con los tres tipos de problemas arriba mencionados, y que estos podrían eventualmente comprometer ciertos principios éticos que gobiernan a los sistemas de inteligencia artificial y vulnerar los derechos humanos, los valores democráticos y el Estado de Derecho (Keats-Citron y Pasquale, 2014; Keats-Citron, 2008 y Kroll et al., 2017). Esto no significa que no existan otros problemas asociados con la recolección masiva de datos y su procesamiento automatizado mediante sistemas de inteligencia. Sin embargo, por razones de formato y extensión, estos serán abordados aquí solo parcialmente.

### Transparencia y responsabilidad algorítmica

La inteligencia artificial, en tanto creación humana, puede reproducir nuestras mayores virtudes y también nuestros peores defectos. En la literatura se señala que los modelos o procesos algorítmicos de toma de decisiones con base en el procesamiento automatizado de datos pueden parecer, en principio, objetivos, racionales, imparciales y libres de todos los prejuicios propios del razonamiento humano (Katyal, 2019). Sin embargo, lo cierto es que los datos y los modelos algorítmicos que los procesan de manera automatizada podrían estar impregnados por los mismos prejuicios irracionales de sus creadores o programadores humanos, bien sea porque no se percatan o tienen en cuenta información sobre prejuicios sistémicos o discriminación estructural, o porque simplemente cometen un error en el diseño o implementación del modelo (Katyal, 2019).

Ello se traduce en problemas de transparencia y responsabilidad algorítmica, cuyo diagnóstico reviste la mayor complejidad dada la naturaleza propietaria de dichos modelos algorítmicos y la dificultad de su análisis o escrutinio público como consecuencia de las restricciones derivadas de la propiedad intelectual (Katyal, 2019 y Levine 2007).

## El control de los sistemas de inteligencia artificial



En la literatura se ha señalado que los problemas relativos al control de los sistemas de inteligencia artificial derivan principalmente de la complejidad de los algoritmos y su arquitectura (Russell, 2019 e Information Society Project, 2017). Sin embargo, el problema del control no se debe confundir con la explicabilidad o inteligibilidad de las decisiones tomadas por dichos sistemas, tema en el cual se profundizará más adelante.

Para Russell (2017), el problema del control de la inteligencia artificial se trata, en esencia, de “(...) cómo asegurar que sistemas con un nivel de inteligencia altamente arbitrario permanezcan estrictamente bajo control humano”. A juicio de este autor, el problema del control implica que podrían existir sistemas de súper inteligencia artificial cuyas acciones sean por definición impredecibles por simples humanos. Tales sistemas podrían operar bajo objetivos establecidos de manera imperfecta e incompleta que riñan con los de las personas y cuya motivación para preservar su propia existencia a fin de alcanzar dichos objetivos sea insuperable (Russell, 2017).

Asimismo, los problemas de control de los sistemas de IA se refieren al grado, naturaleza y extensión de la intervención humana en el proceso automatizado de toma de decisiones. La segunda edición del modelo de marco de gobernanza de la IA propuesto por Singapur (Singapore Digital, InfoComm Media Development Authority y Personal Data Protection Commission, 2020) resulta ilustrativa sobre este punto. Dicho marco de gobernanza se encuentra estructurado sobre una metodología que identifica los riesgos asociados con el funcionamiento de los sistemas de IA y propone una estrategia para identificar los diferentes grados de intervención humana.

En ese marco figura, en primer lugar, el modelo en que el humano interviene activamente y ejerce control absoluto sobre el proceso automatizado de toma de decisiones (*Human-in-the-loop*), en tanto que el sistema de IA se limita a generar recomendaciones o insumos. Ejemplo de ello sería un sistema de IA diseñado para apoyar al médico en el proceso de diagnosticar tal o cual patología, aunque sea él/ella quien tenga la última palabra en materia de diagnóstico. En segundo lugar, el marco adoptado en Singapur describe el modelo en que el humano no interviene en ningún sentido y tampoco tiene control alguno sobre el sistema de IA o sobre sus productos o resultados (*Human-out-of-the-loop*). Un ejemplo de ello sería la elaboración automatizada de perfiles con fines publicitarios. En tercer lugar, en el documento en mención se identifica el modelo en el cual el humano supervisa o monitorea activamente el sistema de IA, de manera que pueda asumir el control en aquellos casos en que este deba afrontar eventos inesperados o indeseables (*Human-over-the-loop* o *Human-on-the-loop*). Ejemplo de ello sería un sistema de IA que calcula la mejor ruta para desplazarse de un punto A a un punto B, en cuyo caso el humano tiene el control de ajustar los parámetros necesarios para completar el trayecto de acuerdo con las circunstancias inesperadas que puedan surgir (Singapore Digital, InfoComm Media Development Authority y Personal Data Protection Commission, 2020).

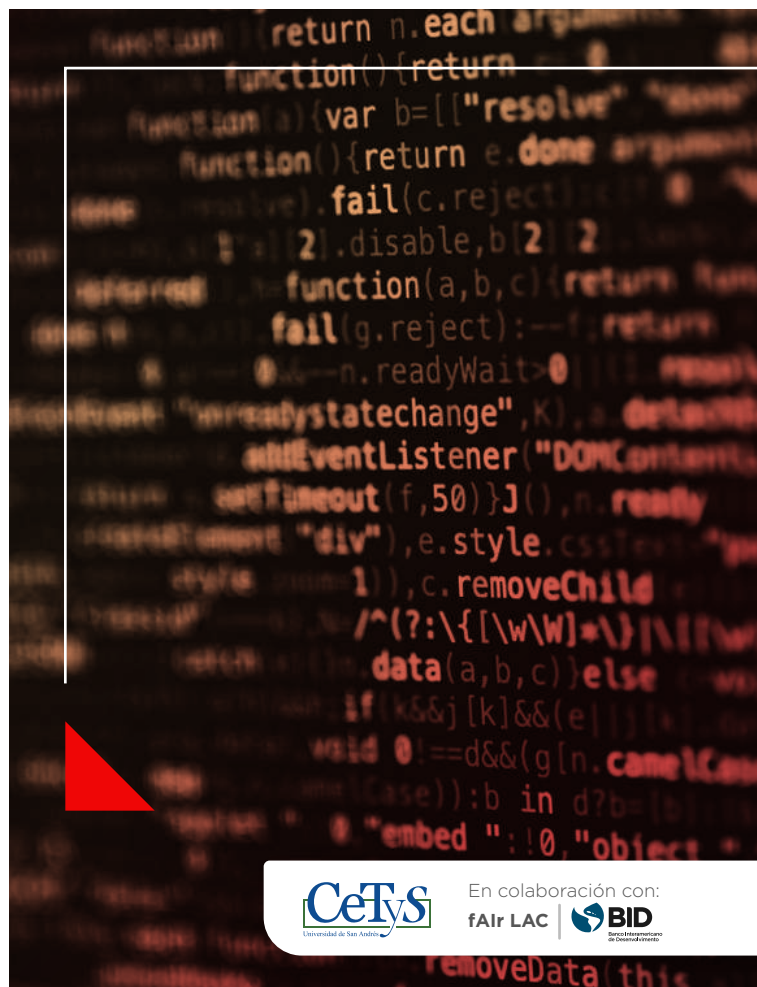


## Explicabilidad e inteligibilidad algorítmica

En la literatura se sostiene que uno de los principales retos del gobierno de la inteligencia artificial estriba en la dificultad técnica de entender, *ex ante* o *ex post*, la manera en que un algoritmo produce tal o cual resultado (The Information Society Project at Yale Law School & Immuta). Igualmente explica que una de las características principales de los métodos del aprendizaje automático (*machine learning*) radica en su capacidad de analizar grandes cantidades de datos e identificar patrones de conducta que incluso pueden ser desconocidos para la mente humana, para luego, con base en este aprendizaje, extrapolar dichos patrones de conducta identificados a situaciones distintas para las cuales no han sido explícitamente programados. Ello implica que los programadores no tendrán la necesidad de incorporar manualmente reglas explícitas para todas y cada una de las situaciones que la máquina pretenda analizar, una labor humana no solamente compleja sino además técnicamente engorrosa y económicamente costosa.

En este contexto, Floridi y Cowls (2019) señalan que la explicabilidad algorítmica reposa sobre dos pilares esenciales. De una parte, esta incorpora el “sentido epistemológico de la ‘inteligibilidad’ (como respuesta a la pregunta ¿cómo funciona?)”, mientras que de otra parte abarca “(...) el sentido ético de la responsabilidad (como respuesta a la pregunta ¿quién es responsable por la manera en que funciona?)” (Floridi y Cowls, 2019). Otros autores (Lehr y Ohm, 2017; Selbst y Barocas, 1918) argumenta que los problemas de explicabilidad de los algoritmos del aprendizaje automático derivan de la naturaleza inescrutable y no intuitiva de las reglas que gobiernan los procesos de toma de decisiones en los cuales se incorpora dicha arquitectura algorítmica.

En este sentido, en la literatura se explica que la naturaleza inescrutable del aprendizaje automático se refiere a “(...) una situación en la que las reglas que gobiernan un proceso de toma de decisiones son tan complejas, numerosas e interdependientes que desafían la inspección práctica y resisten todo intento de comprensión” (Selbst y Barocas, 1918). Asimismo se indica que la naturaleza no intuitiva del aprendizaje automático versa sobre “(...) la inhabilidad de poder estructurar una historia sensible que pueda explicar las relaciones estadísticas del modelo [algorítmico]”. Lo anterior significa que, si bien es posible identificar las relaciones estadísticas que le sirven de base al algoritmo, “(...) dichas relaciones pueden cuestionar las expectativas intuitivas respecto de la relevancia de ciertos criterios que le sirven de base a la decisión [producida por la máquina]” (Selbst y Barocas, 1918).





## 4

## Principales retos éticos y jurídicos de la inteligencia artificial

Todas las situaciones anteriormente descritas sobre la gobernanza de la inteligencia artificial sugieren que su fuerza disruptiva genera una profunda incertidumbre respecto de sus eventuales consecuencias éticas y jurídicas. Por ejemplo, desde una perspectiva ética, Floridi y Cowls (2019) explican que la responsabilidad, transparencia y explicabilidad algorítmica se encuentran estrechamente relacionadas con la protección de los principios éticos de benevolencia, no maleficencia y autonomía que deberían gobernar a los sistemas de inteligencia artificial. En efecto, los citados autores afirman que la única manera de verificar que dichos sistemas harán el bien y preservarán la autonomía en lugar de socavarla, es precisamente mediante la comprensión plena de la manera en que dicho sistema opera y la identificación de quién será el responsable de su funcionamiento indebido (Floridi y Cowls, 2019).

Desde un punto de vista legal, la doctrina ha identificado que los problemas de transparencia, responsabilidad, control y explicabilidad podrían tener consecuencias jurídicas en distintos campos como el mercado laboral, la competencia, la propiedad intelectual, la privacidad, la responsabilidad civil y penal, los derechos humanos y el Estado de Derecho, entre otros (véanse O'Neil, 2016; Rubinstein, Lee y Schwartz, 2008; Zarzky, 2011; OECD, 2019a). Por ejemplo, existen varios puntos de vista y opiniones jurídicas disímiles en torno a la naturaleza jurídica de los datos, su procesamiento, propiedad y control, y sobre la naturaleza, grado y extensión en que dichos datos pueden ser compartidos o transferidos. Por estas razones, la academia y distintas instancias internacionales sostienen que las regulaciones generales de protección de datos personales son insuficientes para garantizar el derecho a la privacidad frente a los retos derivados de su procesamiento automatizado mediante sistemas de inteligencia artificial (véanse Tene y Polonetsky, 2012; Tene y Polonetsky, 2013; Crawford y Schultz, 2014; Katyal, 2019; Solove, 2001; Russell, Dewey y Tegmark, 2015; HM Treasury, 2018).

Asimismo, en la literatura se ha señalado que entre dichos efectos jurídicos figuran, por ejemplo, la perpetuación automatizada de prejuicios sistémicos y discriminación estructural (Katyal, 2019); los problemas en el cumplimiento de los requerimientos de explicabilidad exigidos por el Fair Credit Reporting Act y el Equal Credit Opportunity Act en los Estados Unidos y los impuestos por el GDPR en la Unión Europea (Selbst y Barocas, 2018); la discriminación en el acceso a crédito y al sistema financiero por la

implementación de inteligencia artificial en productos o servicios FinTech (Bartlett et al., 2019); la violación del derecho al debido proceso por parte de sistemas de calificación ejecutados por inteligencia artificial que pueden comprometer el acceso al sistema financiero, laboral o de vivienda (Keats-Citron y Pasquale, 2014), y la discriminación derivada de la aplicación de técnicas de minería de datos e inteligencia artificial para la creación de perfiles que optimicen la lucha contra el terrorismo (Rubinstein, Lee y Schwartz, 2008), entre otros (O'Neil, 2016).

## Inteligencia artificial, inteligencia de datos y privacidad

Aquí se reseñará el marco regulatorio vigente en América Latina con el fin de determinar si existe algún tipo de normatividad que gobierne el diseño, desarrollo e implementación de sistemas de inteligencia artificial, el procesamiento de datos mediante dichos sistemas, y sus posibles consecuencias sobre los derechos humanos, la democracia y el Estado de Derecho. Sin embargo, cabe aclarar que no se tratará de una reseña general y/o exhaustiva, sino limitada al objeto del presente escrito.

Paso seguido se realizará un inventario normativo que permita constatar la existencia o no de normas jurídicas de naturaleza constitucional, legal y reglamentaria que regulen aspectos de la gobernanza de la inteligencia artificial tales como la responsabilidad, transparencia, control y explicabilidad algorítmica, y su relación con los derechos humanos a la privacidad y a la autonomía. A partir de ello se procederá a determinar si actualmente existen normas jurídicas sobre la gobernanza de los sistemas de inteligencia artificial y el procesamiento automatizado de datos personales e institucionales realizado por aquellos, y sobre la toma de decisiones con base en las descripciones, predicciones y prescripciones hechas por máquinas.

Ahora bien, aunque aquí se citarán algunos ejemplos puntuales de regulaciones domésticas para ilustrar los argumentos expuestos, no se pretende realizar un análisis de derecho comparado sobre la regulación de los sistemas de inteligencia artificial en varios países. De allí que la exposición no se enfoque en una nación en particular, sino que, dada la naturaleza normativa de las regulaciones, se parte del supuesto de que estas son compatibles con cualquier Estado de Derecho estructurado bajo un ordenamiento constitucional comprometido con los valores democráticos, el principio de legalidad, la separación de poderes y la protección de los derechos humanos.

Lo primero que se debe precisar es que en América Latina el diseño, desarrollo e implementación de sistemas de inteligencia artificial no opera dentro de un vacío jurídico. Al igual que en el caso de cualquier actividad científica, industrial o comercial, los sistemas de inteligencia artificial se encuentran sometidos a una variedad de regulaciones de distinta naturaleza, entre ellas las atinentes a los principios democráticos del Estado de Derecho; los derechos humanos; el régimen de intercambio de mensajes de datos; la neutralidad tecnológica; la propiedad intelectual; la protección del consumidor; la competencia; la protección y procesamiento de datos personales o institucionales; y el derecho tributario, el derecho penal y la responsabilidad civil, entre otras.

Como ya se ha indicado en este documento, el procesamiento automatizado de datos personales e institucionales mediante sistemas de inteligencia artificial para describir, predecir o prescribir información suscita muchos interrogantes relativos al régimen jurídico aplicable a dichos datos para proteger la privacidad. Bajo esta premisa, a continuación se describirán las reformas más recientes introducidas al régimen general de protección de datos en Europa, para luego, con base en dichas regulaciones, analizar la manera en que algunos países latinoamericanos han incorporado recientemente o están en vías de incorporar regulaciones similares en sus ordenamientos jurídicos internos.

## El derecho a obtener explicaciones sobre decisiones individuales automatizadas (*right to explanation*) en Europa

El Reglamento General de Protección de Datos (RGPD (<https://gdpr.eu/>)) europeo constituye una referencia obligada en materia de protección de datos, ya que contiene regulaciones específicas sobre el procesamiento automatizado de datos personales. Puntualmente, la Sección 4 del Artículo 22 dispone lo siguiente, en materia de “Decisiones individuales automatizadas, incluida la elaboración de perfiles:

1. *Todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado –inclusive la elaboración de perfiles-- que produzca efectos jurídicos en él [o ella] o [que] le afecte significativamente de modo similar.*
2. *El apartado 1 no se aplicará si la decisión...:*
  - a. *...es necesaria para la celebración o la ejecución de un contrato entre el interesado y un responsable del tratamiento;*
  - b. *...está autorizada por el Derecho de la Unión o de los Estados miembros que se aplique al responsable del tratamiento y que establezca asimismo medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado, o*
  - c. *se basa en el consentimiento explícito del interesado.*
3. *En los casos a los que se refiere el apartado 2, literales (a) y (c), el responsable del tratamiento adoptará las medidas adecuadas para salvaguardar los derechos y libertades, y los intereses legítimos del interesado, como mínimo el derecho a obtener intervención humana por parte del responsable, a expresar su punto de vista y a impugnar la decisión.*
4. *Las decisiones a las que se refiere el apartado 2 no se basarán en las categorías especiales de datos personales contempladas en el artículo 9, apartado 1, salvo que se aplique el artículo 9, apartado 2, literales (a) o (g), y que se hayan tomado medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado”.*





Mucho se ha debatido sobre la existencia o no de un derecho a obtener una explicación sobre una decisión individual automatizada (*right to explanation*) en el RGPD, aplicable a los casos en que los datos personales sean procesados automáticamente por un sistema de inteligencia artificial sin intervención humana. En un primer momento, Goodman y Flaxman (2017) argumentaron que el RGPD creó un “derecho a obtener una explicación” cuando una decisión se toma de manera automatizada con base en el procesamiento de datos personales. Sin embargo, esos autores no profundizaron sobre los fundamentos, naturaleza y estructura de dicho derecho a obtener una explicación.

De otra parte, Wachter, Mittelstadt y Floridi (2017) negaron la existencia del derecho a obtener una explicación en el RGPD, dada la ambigüedad del texto de dicho instrumento normativo. En su lugar, con base en una interpretación de los artículos 13, 14, 15 y 22 del RGPD, los citados autores concluyen que existe un “derecho a ser informado” acerca de la funcionalidad de un sistema de inteligencia artificial que procesa datos personales de manera automatizada para producir una decisión sin intervención humana. Las conclusiones de Wachter, Mittelstadt y Floridi (2017) deben ser analizadas en contexto, a lo cual puede contribuir la clasificación de las diferentes explicaciones que un individuo puede solicitar respecto del funcionamiento de un sistema de inteligencia artificial.

Dichas explicaciones pueden ser de dos clases: (i) sobre la funcionalidad del sistema y (ii) sobre decisiones específicas. De una parte, para los autores aludidos la funcionalidad del sistema comprende la “(...) lógica, significado, consecuencias previstas, funcionalidad general de un sistema automático de toma de decisiones tales como sus especificaciones técnicas, árboles de decisión, modelos predefinidos, criterios [y] estructuras de las clasificaciones (...)”. De otra parte, las explicaciones acerca de decisiones específicas pueden recaer sobre “(...) el razonamiento, razones y circunstancias individuales de una decisión específica adoptada de manera automática, tales como los factores de ponderación, las reglas de la decisión generadas para un caso específico por una máquina, [la] información de referencia o [los] grupos de perfiles”.

Wachter, Mittelstadt y Floridi (2017) también proponen una clasificación adicional sobre el momento en que se materializan las explicaciones y el proceso automatizado de toma de decisiones. Para estos autores, las explicaciones son *ex ante* cuando se presentan con anterioridad a aquel, por lo que estas únicamente podrán versar sobre la funcionalidad del sistema y no sobre la decisión específica. En cambio, las explicaciones son *ex post* cuando se presenten con posterioridad a la toma de la decisión individual automatizada, y en tal sentido podrían referirse tanto a la funcionalidad del sistema como a la decisión específica.

Para entender el alcance de este debate sobre la existencia de un derecho a obtener una explicación en el RGPD, vale la pena presentar la posición adoptada por Selbst y Powles (2017). Estos autores indican que no existe ninguna disposición en el RGPD que establezca con claridad un derecho a obtener una explicación sobre las decisiones individuales automatizadas. Sin embargo, sí argumentan que existe un derecho a obtener “información significativa sobre la lógica aplicada” en la decisión individual automatizada de acuerdo con lo previsto en el artículo 13 (2)(f). Ello se desprende de una interpretación funcional y flexible de los artículos 13, 14 y 15 del RGPD.

Si bien las posturas hasta ahora expuestas sobre la existencia de un derecho a obtener explicaciones sobre decisiones automatizadas son razonables, todo parece indicar que la controversia interpretativa fue zanjada a partir de los lineamientos sobre decisiones automatizadas individuales y perfilamiento expedidos por el Grupo de Trabajo del Artículo 29 el 6 de febrero de 2018<sup>2</sup>. A la luz de dichos lineamientos, así como de una extensa interpretación de los artículos 13, 14, 15 y 22 del RGPD y sus considerandos, Margot Kaminski (2019) sostiene que no existe duda sobre la existencia de un derecho a obtener explicaciones acerca de una decisión individual automatizada tomada como medida de responsabilidad algorítmica.

El ejercicio de tal derecho merece algunas precisiones. De acuerdo con el artículo 22 del RGPD, la literatura y el Grupo de Trabajo del Artículo 29 han señalado que el derecho a obtener explicaciones sobre decisiones individuales automatizadas solamente puede ser ejercido cuando la decisión se encuentre “únicamente basada en el tratamiento automatizado, incluida la elaboración de perfiles” que produzca “efectos jurídicos” o “le afecte significativamente de modo similar”.

Frente al primer requisito en mención, el Grupo de Trabajo del Artículo 29 señaló que el derecho conferido por el artículo 22 del RGPD únicamente podrá ser ejercido cuando se trate de una decisión que cause efectos “seriamente significativos” y estableció unos criterios que sirven para determinar lo que ello significa. Como ejemplo, el Grupo de Trabajo señala que entre dichas decisiones figuran aquellas que afecten situaciones financieras, el acceso a la salud o a la educación, que nieguen el acceso a un empleo o que pongan a una persona en una situación de “seria desventaja”<sup>3</sup>.

En lo relativo al segundo requisito sobre la intervención humana, el Grupo de Trabajo indicó que para que la decisión quede sometida al artículo 22 se requiere que la intervención humana sea significativa en el sentido de que debe ser “realizada por alguien con la autoridad y competencia para cambiar la decisión”; para ello será necesario que tenga acceso a información que vaya más allá de los productos o resultados que arroje el algoritmo<sup>4</sup>.

<sup>2</sup> Véase el Article 29 Data Protection Working Party, Guidelines on Automated Individual Decision-Making And Profiling For The Purposes Of Regulation 2016/679, 17/EN. WP 251rev.01 (Feb. 6, 2018).

<sup>3</sup> Véase el Article 29, *op cit.*

<sup>4</sup> *Ibid.*

## El derecho a obtener explicaciones sobre decisiones individuales automatizadas en América Latina

Aquí se analiza si el derecho a obtener explicaciones sobre decisiones individuales automatizadas existe en América Latina. Para ello se estudiarán la Lei Geral de Proteção de Dados Pessoais de Brasil (LGPD) de 2018, el Proyecto de Reforma del Régimen de Protección de Datos Personales en Argentina y el Proyecto de Reforma de Ley de Protección de Datos de Chile.

En Brasil, la Lei Geral de Proteção de Dados Pessoais (LGPD) de 2018 (<https://lcpd.com.br/home-eng>) se encuentra estructurada a partir del RGPD y del California Consumer Privacy Act de 2018. Puntualmente, su artículo 20 se refiere al procesamiento automatizado de los datos personales:

*“El interesado tiene derecho a solicitar la revisión de las decisiones adoptadas únicamente sobre la base del tratamiento automatizado de datos personales que afecten a sus intereses, incluidas las decisiones destinadas a definir su perfil personal, profesional, de consumo y de crédito o aspectos de su personalidad”.*

De una simple lectura de esta disposición normativa se desprende la existencia de un derecho a obtener explicaciones sobre decisiones individuales automatizadas. En efecto, los párrafos 1º y 2º del mismo artículo indican que el responsable, a solicitud del titular, debe informar sobre los criterios y procedimientos utilizados para la toma de una decisión automatizada. En caso de que el responsable o encargado aluda a la protección de secretos industriales para negarse a brindar información, queda la posibilidad de solicitar una auditoría a la autoridad nacional.

En Argentina, el Proyecto de Reforma del Régimen de Protección de Datos Personales también se refirió al tratamiento automatizado de tales datos para la producción de decisiones individualizadas<sup>5</sup>. El artículo 28 del citado Proyecto de Ley definió los derechos del titular sobre decisiones automatizadas en el contexto del contenido de la información:

<sup>5</sup> El Proyecto de Ley se puede consultar en: [https://www.argentina.gob.ar/sites/default/files/mensaje\\_nde\\_147-2018\\_datos\\_personales.pdf](https://www.argentina.gob.ar/sites/default/files/mensaje_nde_147-2018_datos_personales.pdf)



*“Contenido de la información. La información debe ser suministrada en forma clara, exenta de codificaciones y, en su caso, acompañada de una explicación de los términos que se utilicen, en lenguaje accesible al conocimiento medio de la población, y debe versar sobre:*

*[...] (h) La existencia de decisiones automatizadas, incluida la elaboración de perfiles a que se refiere el artículo 32 y, al menos en tales casos, información significativa sobre la lógica aplicada, sin que ello afecte derechos intelectuales del responsable del tratamiento”.*

Por su parte, el artículo 32 del citado Proyecto de Ley dispone lo siguiente:

*“ARTÍCULO 32.- Valoraciones personales automatizadas. El titular de los datos tiene derecho a oponerse a ser objeto de una decisión basada únicamente en el tratamiento automatizado de datos, incluida la elaboración de perfiles, que le produzca efectos jurídicos perniciosos o lo afecte significativamente de forma negativa.*

*El titular de los datos no podrá ejercer este derecho si la decisión:*

- (a) Es necesaria para la celebración o la ejecución de un contrato entre el titular de los datos y el responsable del tratamiento;*
- (b) Está autorizada por Ley;*
- (c) Se basa en su consentimiento expreso.*

*En los casos a que se refieren los incisos (a) y (c), el responsable del tratamiento debe adoptar las medidas adecuadas para salvaguardar los derechos del titular de los datos”.*

Cabe notar que la redacción de los citados artículos 28 y 32 del Proyecto de Ley guardan una estrecha relación con los artículos 13, 14, 15 y 22 del RGPD. En efecto, la redacción del artículo 28 es similar a la del artículo 22 del RGPD, razón por la que se deduce que contiene los pilares estructurales del derecho a obtener explicaciones sobre decisiones individuales automatizadas.

En Chile, por su parte, el Proyecto de Reforma del Régimen de Protección de Datos<sup>6</sup> contiene varias disposiciones normativas que permiten establecer la existencia del derecho a obtener explicaciones sobre las decisiones individuales automatizadas. De una parte, el artículo 8 del citado proyecto de reforma dispone:

**ff** “ARTÍCULO 8.- Derecho de oposición. El titular de datos tiene derecho a oponerse ante el responsable a que se realice un tratamiento específico o determinado de los datos personales que le conciernan, en los siguientes casos:  
(...)  
(c) Cuando se realice tratamiento automatizado de sus datos personales y se adopten decisiones que impliquen una valoración, evaluación o predicción de su comportamiento realizada únicamente en base a este tipo de tratamiento, salvo las excepciones previstas en el artículo 15 ter de esta ley”.

En este sentido, el artículo 15 ter. del referido proyecto de ley señala:

**ff** “Artículo 15 ter. - Tratamiento automatizado de grandes volúmenes de datos.- El responsable de datos puede establecer procedimientos automatizados de tratamiento y de transferencia de grandes volúmenes de datos, siempre que estos cautelen los derechos del titular y el tratamiento guarde relación con las finalidades de las personas o entidades participantes.

*El titular de datos tiene derecho a solicitar al responsable que ninguna decisión que le **afecte de manera significativa** (énfasis propio) se adopte exclusivamente basada en el tratamiento automatizado de sus datos, salvo que sea necesario para la celebración o ejecución de un contrato entre el titular y el responsable, exista consentimiento previo y explícito del titular o lo disponga la ley”.*

<sup>6</sup> El proyecto de ley se puede consultar en: <http://www.informatica-juridica.com/proyecto-de-ley-y/proyecto-ley-proteccion-datos-chile-abril-2017/>

Cabe señalar que la redacción de los artículos del proyecto de reforma de Chile dista de la de los artículos 13, 14, 15 y 22 del RGPD europeo. En efecto, el artículo 15 ter. parece contener más un derecho de oposición a la producción de decisiones individuales automatizadas que causen una afectación significativa, que un derecho a obtener explicaciones sobre dicha decisión automatizada. A diferencia del RGPD, así como de la nueva

ley de protección de datos de Brasil y del proyecto de reforma de Argentina, el proyecto de reforma del régimen de protección de datos de Chile no contiene aparentemente el derecho a conocer la lógica aplicada en el tratamiento automatizado de datos para la producción de una decisión individual. Sin embargo, habrá que esperar a la redacción final del proyecto de ley que apruebe la Legislatura chilena.

De cualquier manera, a los proyectos que cursan trámite legislativo en Argentina y Chile todavía les queda un largo camino por recorrer. Una vez aprobados los textos finales, procederá estudiar la manera en que las agencias de protección de datos (DPA por sus siglas en inglés) brasileña, argentina y chilena interpretarán estas disposiciones a efectos de determinar si seguirán de cerca la interpretación del artículo 22 del RGPD hecha por el Grupo de Trabajo del artículo 29, o si allí se realizará una interpretación distinta que restrinja o amplíe la responsabilidad algorítmica derivada del procesamiento automatizado de datos por parte de sistemas de inteligencia artificial.

El desarrollo, diseño e implementación de servicios digitales y/o modelos de negocio que incorporen el procesamiento automatizado de datos personales mediante sistemas de IA debe ajustarse plenamente a la Constitución y leyes vigentes. Bajo esta óptica, el diseño e implementación de dichos sistemas deben estar sometidos al bloque de legalidad entendido en su dimensión más amplia, es decir, un conjunto de normas de naturaleza convencional, constitucional, legal, reglamentaria y regulatoria.

De igual forma, la ética digital está desempeñando un papel determinante en la interpretación y actualización de la regulación vigente, al tiempo que permite a los sectores público y privado dar un paso más allá de la legalidad para proteger el derecho que tienen todos los individuos a tomar sus propias decisiones de manera autónoma, informada y sin injerencias indebidas, no consentidas o ilegales.

Como se indicó al principio de este documento, el Grupo de Expertos de Alto Nivel de la Unión Europea en Inteligencia Artificial (HL-Expert Group o HLEG) publicó recientemente unos lineamientos éticos (<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>) para promover el diseño, desarrollo e implementación de sistemas de inteligencia artificial confiables. Estos lineamientos de la UE reposan sobre la premisa de que los sistemas de inteligencia artificial confiables deben ser: (i) legales, (ii) éticos y (iii) técnicamente sólidos.

El citado informe deja en claro desde el inicio que los sistemas de inteligencia artificial deben ser legales en el sentido de que deben cumplir a cabalidad con todas las leyes y regulaciones existentes. Sin embargo, en el informe no se señalan los estándares regulatorios mínimos a los que se encuentra sometido el diseño, desarrollo e implementación de los sistemas de IA. Tampoco se sugiere algún tipo de estrategia o método para que dichos sistemas cumplan en forma



debida con las regulaciones aplicables. Una vez hecha esta importante aclaración, el informe es enfático en señalar que se ocupará únicamente de los lineamientos correspondientes a los componentes éticos y técnicos sobre los que debe gravitar todo sistema de IA confiable.

El HLEG sostiene que la IA confiable no se puede lograr únicamente mediante el cumplimiento de la ley, sino que es necesario garantizar que estos sistemas se sometan a unos principios y valores éticos. Lo anterior se da como consecuencia de que, a juicio del HLEG, las leyes pueden estar desactualizadas frente a los retos que impone el vertiginoso desarrollo de nuevas tecnologías, no ser adecuadas para resolver ciertos problemas o marchar a un compás distinto al de las normas éticas<sup>7</sup>.

El HLEG también indica que los sistemas de IA confiables deben ser sólidos en términos técnicos y sociales para que los individuos y la sociedad tengan la certeza de que dichos sistemas no causarán daños indeseados. Para ello, el grupo de expertos sugiere que los sistemas de IA deben operar en una forma segura, confiable y bajo ciertas salvaguardas que prevengan cualquier tipo de impactos adversos en los individuos y en la sociedad.

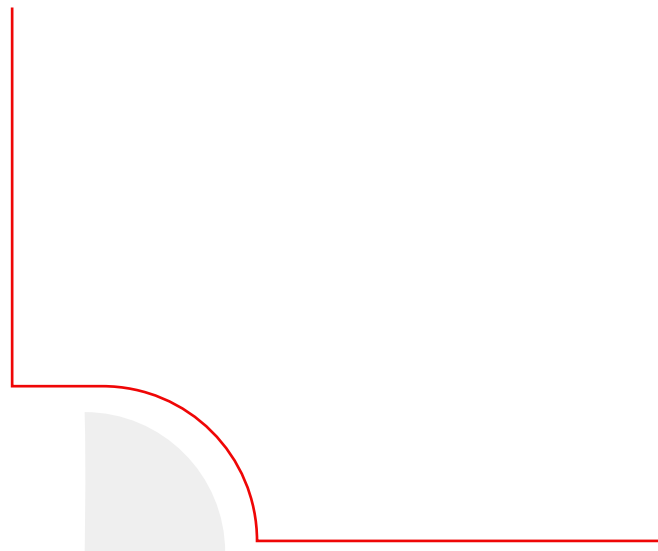
El HLEG sostiene que la inteligencia artificial ética debe “(...) ayudar a las personas a elegir mejor y con más conocimiento de causa en función de sus objetivos. Debe actuar como facilitadora de una sociedad floreciente y equitativa, apoyando la intervención humana y los derechos fundamentales sin disminuir, limitar o desorientar la autonomía humana” (HLEG, 2019). Para garantizar el cumplimiento de estos ideales, el HLEG de la UE señala que la inteligencia artificial fiable debe reposar sobre cuatro principios éticos: (i) respeto de la autonomía humana, (ii) prevención del daño, (iii) justicia y (iv) explicabilidad (HLEG, 2019).

<sup>7</sup> Véase p. 7 del informe de lineamientos éticos de la UE



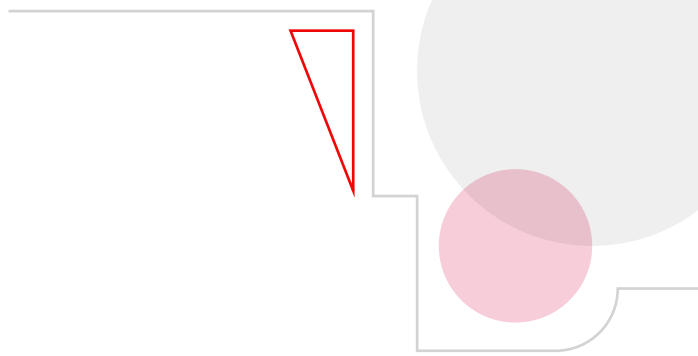
En el mismo sentido, en las recomendaciones del Consejo de Inteligencia Artificial de la OCDE (OECD, 2019a) se señala que el desarrollo e implementación en los países adherentes debe respetar el Estado de Derecho, los derechos humanos y los valores democráticos. Para ello, el desarrollo e implementación de sistemas de inteligencia artificial debe respetar la libertad, la dignidad humana, la privacidad y la protección de datos, así como la igualdad, además de prevenir la discriminación y promover la diversidad, la justicia social y los derechos laborales.

En este contexto, la OCDE (OECD, 2019a) recomienda que el proceso de desarrollo e implementación de sistemas de inteligencia artificial persiga fines beneficiosos para las personas y el planeta, expanda las capacidades humanas e impulse la creatividad, promueva la inclusión de las poblaciones marginadas, reduzca las desigualdades económicas, sociales y de género, y proteja el medio ambiente.



En este orden de ideas, la Unión Europea y la OCDE han sostenido que el proceso de diseño, desarrollo e implementación de sistemas de inteligencia artificial debe girar en torno a cuatro principios éticos: (i) transparencia, (ii) responsabilidad, (iii) control y (iv) explicabilidad. Con base en dichas recomendaciones, se sugiere que las entidades contratantes realicen todos los estudios y análisis necesarios para garantizar que los bienes o servicios digitales adquiridos para la gestión o prestación de una actividad o servicio a cargo de una autoridad administrativa observen dichos principios legales y éticos. Lo anterior, por supuesto, como requisito adicional al estricto cumplimiento de los principios democráticos del Estado de Derecho.

Sin embargo, esto no se logra únicamente con el diseño de procedimientos, reglas y protocolos en el papel. Todo Estado de Derecho digno de su nombre, así como los principios democráticos sobre los que este descansa, requieren el diseño e implementación de herramientas o métodos **X-por-diseño (X= legalidad, privacidad o ética digital) y por defecto** para integrar la legalidad, la privacidad y la ética digital dentro de las tecnologías, operaciones y arquitectura de los bienes o servicios digitales y/o modelos de negocio que incorporen el procesamiento automatizado de datos personales mediante sistemas de IA<sup>8</sup>.



<sup>8</sup> Esta sección recoge en buena parte las ideas que el autor ha presentado en otros escenarios (Castaño, 2019b). Con fundamento en este estudio jurídico elaborado por el autor, la Alcaldía Mayor de Bogotá expidió la Directiva Distrital No. 10 de 17 de diciembre de 2019, la cual es únicamente vinculante para las autoridades del Distrito Capital.

En efecto, en la literatura se explica que las herramientas o métodos X-por-diseño y por defecto anteriormente aludidos no se añaden después del hecho o como una adición posterior (*add-on*) al diseño, desarrollo e implementación del bien o servicio digital (Cavoukian, S.f.). La principal consecuencia de ello es que la legalidad, privacidad y ética digital se convierten en un componente esencial y por defecto de la funcionalidad del producto o servicio digital prestado por un agente público o privado. En consecuencia, tal componente debería hacer parte integral del bien o servicio digital, pero sin que ello comprometa su funcionalidad o aptitud para cumplir con la finalidad que estos persiguen (Cavoukian, S.f.).

Es así como la legalidad, privacidad y ética digital deberían estar plenamente integradas en las tecnologías, operaciones y arquitecturas de los bienes y servicios digitales de una manera holística, integradora y creativa. A la luz de los trabajos y propuestas de Cavoukian (S.f.), aquí se sugiere que las herramientas o métodos X-por-diseño deben ser:

- ▶ **Holísticos:** El diseño, desarrollo e implementación de herramientas y arquitecturas de X-por-diseño deben considerar todas las visiones, contextos y circunstancias derivadas de diferentes disciplinas.
- ▶ **Integradores:** La tecnología es un espacio donde deben concurrir todas las visiones e intereses presentes en una comunidad democrática pluralista. Se recomienda que todos los actores interesados sean consultados en lo que respecta al diseño, desarrollo e implementación de dichas herramientas o métodos con el fin de lograr un balance entre los diferentes intereses presentes en el procesamiento automatizado de datos personales mediante sistemas de IA.
- ▶ **Creativos:** El diseño, desarrollo e implementación de herramientas o arquitecturas de X-por-diseño deben ser creativos, en la medida en que la plena integración de los valores por diseño (legalidad, privacidad o ética digital) a veces requiere que se reinventen las opciones técnicas existentes; esto por cuanto hay ocasiones en las que las alternativas vigentes son inaceptables o se quedan cortas para cumplir tal cometido. Por esta razón, el diseño, desarrollo e implementación de herramientas o arquitecturas de X-por-diseño debería estar siempre a la vanguardia.

## 6

## □ iniciativas de experimentación regulatoria y ética

En este orden de ideas, la estructuración y puesta en marcha de iniciativas de experimentación regulatoria (*regulatory sandboxes*) podrían ser determinantes en la interpretación, implementación y actualización de las regulaciones vigentes en materia de protección de datos personales, en la medida en que permitiría a los DPA y a los demás actores interesados descubrir, mediante la innovación y experimentación regulatorias, cuál sería la mejor forma de regular el derecho a obtener explicaciones sobre decisiones individuales automatizadas sin comprometer la innovación.

Estas iniciativas son el escenario predilecto para el diseño, desarrollo e implementación de herramientas y arquitecturas digitales X-por-diseño, dado que facilitan la experimentación más allá del cumplimiento de la letra y espíritu de la ley al permitir que los actores públicos y privados introduzcan innovaciones en las formas de proteger los derechos fundamentales a la privacidad, el debido proceso y la autonomía personal. Ello se puede lograr no solo mediante autorizaciones para probar tecnologías y modelos de negocio que incorporen el procesamiento automatizado de datos personales mediante sistemas de IA sin estar sometidos a las regulaciones vigentes, sino también creando y ensayando regulaciones experimentales para medir su impacto real en dichos servicios o modelos de negocio.

Cabe recordar aquí que estas iniciativas experimentales hacen referencia a una zona de ensayos de propuestas regulatorias que cuenta con una estructura de participación abierta, en donde el sector público, las empresas y los ciudadanos definen cuál es el mejor enfoque reglamentario sobre determinada tecnología. Se busca propiciar un entorno seguro para la experimentación, recolección de experiencias y pruebas de nuevas tecnologías o modelos de negocio que entrañan un riesgo sistémico sin que este ejercicio sea obstaculizado por la rigidez jurídica existente.

Bajo este esquema, los privados no quedan sujetos a multas y/o sanciones, ni se los hace responsables en el sentido tradicional a medida que se van definiendo escenarios de propuestas reguladoras. De esta forma, es posible crear nuevas reglas y facilitar el acceso de los consumidores en un ambiente controlado, lo que permite flexibilizar o definir posteriormente la política pública en el mundo real (Jiménez y Hagan, 2019).



La noción de iniciativa de experimentación o sandbox<sup>9</sup> intenta redefinir la forma en el que el Estado participa en la elaboración de la política pública; ya no se trata de hablar de gobernabilidad sino de gobernanza, pues allí la toma de decisiones entraña procesos abiertos de participación y experimentación que promueven la retroalimentación de varios sectores de la sociedad.

Así las cosas, la implementación del esquema propuesto debe hacerse por etapas y teniendo en cuenta que se trata de un área de experimentación donde se elabora un prototipo de regulación. Los aspectos que se deben tener en cuenta para comenzar esta iniciativa son: (i) objetivos, sector y propósito de la zona de experimentación; (ii) pautas que determinen quiénes participan en la iniciativa y cuáles serán los requisitos habilitantes para hacerlo; (iii) restricciones de acceso a la zona de experimentación; (iv) naturaleza del acto jurídico que sirve de carta de exención reglamentaria, y autoridad competente y sus límites; (v) evaluación de los resultados de la experimentación y posibilidades de variar los parámetros ya establecidos; (vi) consentimiento informado de los agentes y usuarios participantes.

Las principales propuestas de iniciativas de experimentación regulatoria han tenido lugar en el sector financiero, probablemente a partir de que el Banco Mundial reconociera el hecho de que “han fracasado los intentos de desarrollo basados en el protagonismo del Estado, pero también fracasarán los que se quieran realizar a sus espaldas” (Banco Mundial, 1997).

El primer entorno de experimentación regulatoria tuvo lugar en el sector financiero en el Reino Unido, donde participaron compañías de gestión y consumo, bancos, seguros y cadenas de bloques (blockchains). Durante el primer año se cumplieron los objetivos y ahora las empresas son más competitivas en un mercado más amplio (Financial Conduct Authority, 2015). El informe sobre las lecciones aprendidas muestra que el proyecto permitió una innovación significativa en el sector financiero, mayor volumen de inversión y un crecimiento exponencial de la tecnología (Financial Conduct Authority, 2016).

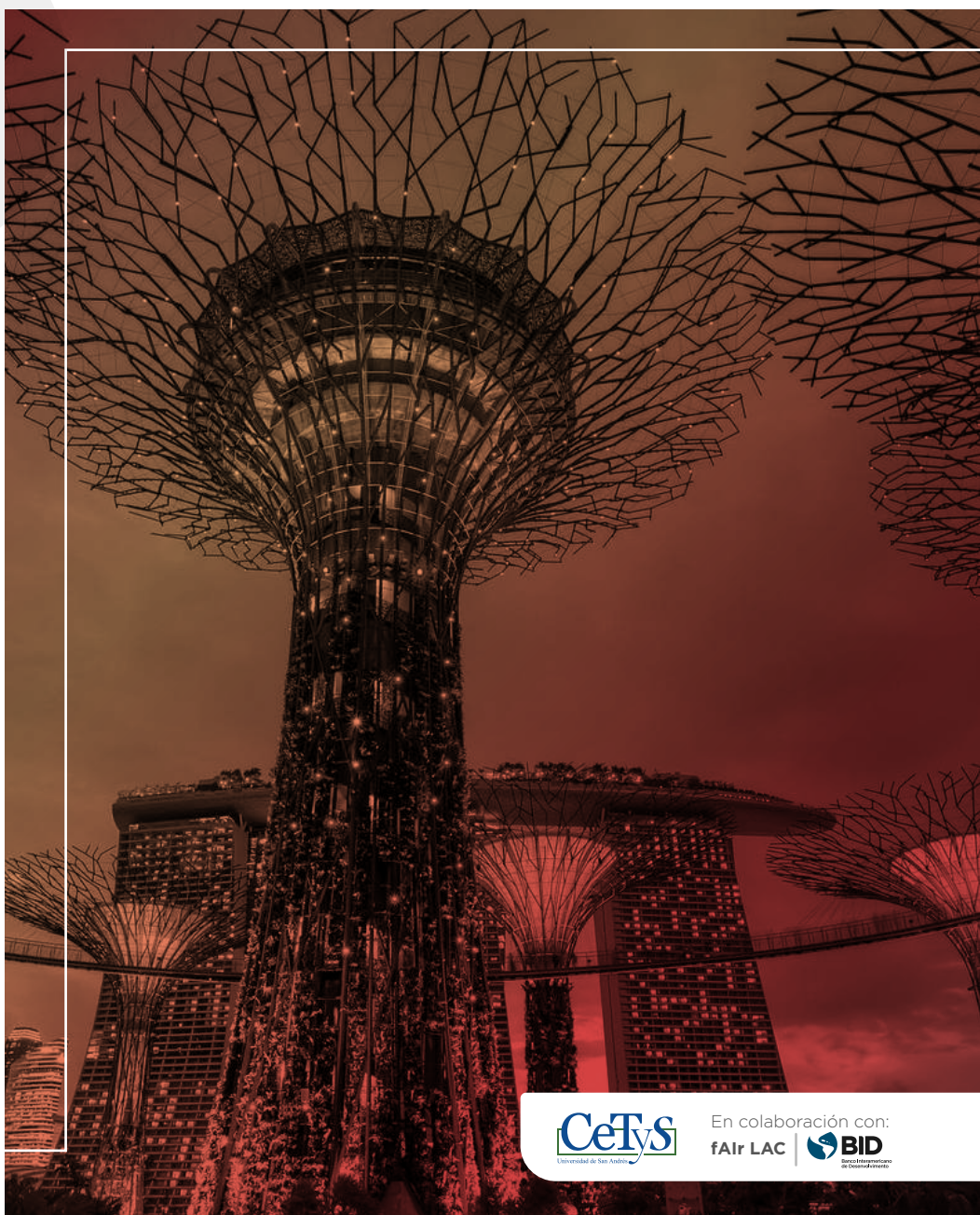
Asimismo, Abu Dabi decidió abrir un Laboratorio de Regulación (RegLab) para su mercado global para que los participantes de FinTech pudieran desarrollar y probar su propuesta por un período de hasta dos años (Abu Dhabi Global Market, 2016).

Estados Unidos siguió la tendencia con una de las primeras iniciativas de experimentación regulatoria en el continente, el Arizona Regulatory Sandbox Program, a través del cual se abordaron tres categorías FinTech: (i) transferencias electrónicas de dinero, (ii) préstamos al consumidor y (iii)

<sup>9</sup> *“Arenera donde los niños juegan, construyen y experimentan. Por extensión, ‘entorno de experimentación’.”*

asesoramiento de inversión. Aquí la zona de experimentación permite la movilidad en estas áreas, de forma que las regulaciones que estén por fuera su entorno de ensayo permanecen intactas. Su vigencia es por un periodo de dos años, y la propuesta de negocio en la que el uso de la tecnología puede ser problemática debe ser nueva o emergente (Watkins, Daniels y Slayton, 2018). Con ello se buscaba aprovechar el potencial de este entorno experimental para mitigar las distorsiones del mercado creadas por las barreras regulatorias.

Otro ejemplo exitoso se da en Singapur, en donde cualquier empresa que utilice tecnología de forma innovadora en los servicios financieros puede acceder al Centro Financiero Inteligente, un ecosistema regulatorio controlado en donde las normas y requisitos reglamentarios prescritos por la Autoridad Monetaria son más laxos (Monetary Authority of Singapore, 2018). A partir de esta idea, el Estado se convierte en facilitador de la elaboración de normas que cuenten con la participación de todos los actores interesados en el desarrollo de nuevas tecnologías, modelos de negocio y de los instrumentos regulatorios apropiados para impulsar la innovación en distintos sectores.





## Conclusiones y recomendaciones

En el presente documento se identificó, con apoyo en la literatura y en los instrumentos internacionales adoptados recientemente, el concepto de inteligencia artificial, su fuerza disruptiva potencial, y sus principales problemas de gobernanza relacionados con la transparencia, responsabilidad, control y explicabilidad algorítmica.

Asimismo, se propuso que la privacidad, agencia, autonomía y autodeterminación son derechos fundamentales que se encuentran en una intersección donde confluyen principios éticos y normas jurídicas, y con los cuales se busca proteger un ámbito de privacidad y autonomía personal, cumpliendo un doble propósito: proteger nuestra información personal, así como la forma, grado y extensión en que decidimos compartirla con otras personas, y salvaguardar la manera en que tomamos nuestras decisiones personales de manera libre y autónoma con base en dicha información y sin ningún tipo de injerencia externa no consentida, indebida o ilegal.

En el contexto internacional, el derecho a obtener una explicación de la decisión tomada con base en el procesamiento automatizado de datos personales por sistemas de IA constituye un primer paso importante para proteger los derechos fundamentales antes mencionados. Sin embargo, a la plena articulación e implementación de este derecho en América Latina todavía le queda un largo camino, a juzgar por los proyectos de reforma que actualmente cursan el trámite legislativo en Argentina y Chile. Una vez aprobados los textos finales, se recomienda estudiar la manera en que los DPA brasileño, argentino y chileno interpretan estas disposiciones a efectos de determinar si siguen de cerca la lectura del artículo 22 del RGPD hecha por el Grupo de Trabajo del artículo 29, o si realizan una interpretación distinta que restrinja o amplíe la responsabilidad algorítmica derivada del procesamiento automatizado de datos por parte de sistemas de inteligencia artificial.

Aunado a lo anterior, se explicaron las razones por las que el derecho a obtener una explicación se podría quedar eventualmente corto para proteger nuestra información personal, la forma, grado y extensión en que decidimos compartirla con otros, y la manera en que tomamos nuestras decisiones personales de manera libre y autónoma con base en dicha información y sin ningún tipo de injerencia externa no consentida, indebida o ilegal. Por estos motivos, se argumentó que la ética digital podría desempeñar un papel crítico para interpretar, ajustar y actualizar las regulaciones vigentes.

Para ello se recomienda la creación y puesta en marcha de iniciativas de experimentación regulatoria, toda vez que estas constituyen el escenario ideal para el diseño, desarrollo e implementación de herramientas y arquitecturas digitales “X-por-diseño”. Tales ejercicios permiten la experimentación más allá del cumplimiento de la letra y espíritu de la ley, ya que los actores públicos y privados pueden crear nuevas formas de proteger los derechos fundamentales a la privacidad, el debido proceso y la autonomía personal.

Cabe señalar, sin embargo, que la eventual proliferación de este tipo de iniciativas en los países de América Latina podría devenir en la “balcanización tecnológica” del continente, es decir, en una suerte de fragmentación regulatoria derivada de los diferentes regímenes jurídicos y éticos que puedan adoptar los países de la región de manera descoordinada, en perjuicio del emprendimiento y la innovación.

En tal sentido, lo razonable sería optar por la creación coordinada de iniciativas regulatorias experimentales que promuevan la innovación, el emprendimiento y la regulación uniforme de la tecnología aplicable a modelos de negocio que reposen sobre el procesamiento automatizado de datos personales por sistemas de IA. A ello pueden contribuir también las instancias multilaterales, considerando la posibilidad de incluir cláusulas o disposiciones contractuales que promuevan prácticas en materia de métodos o herramientas X-por diseño y por defecto (X=Legalidad, Privacidad o Ética Digital). Dentro del contexto latinoamericano, la iniciativa **fAIR Lac del BID** está llamada a desempeñar un papel preponderante armonizando y coordinando las iniciativas de experimentación o arenas regulatorias, y también documentando y compartiendo las experiencias y enseñanzas derivadas de ellas.

En suma, los entornos regulatorios y éticos experimentales constituyen el espacio privilegiado donde deberían concurrir activamente todos los actores interesados para forjar de manera coordinada la regulación que gobierne a la Cuarta Revolución Industrial en todos sus aspectos. Pero ello no ocurrirá únicamente con la expedición de normas en el papel, sino que estas deben ser llevadas a la práctica mediante herramientas o métodos regulatorios alternativos que tengan un impacto real en la protección del derecho fundamental del individuo a tomar decisiones informadas y sin ningún tipo de injerencia externa ilegal, indebida o no consentida.

## Referencias bibliográficas

Abu Dhabi Global Market. 2016. *ADMG Launches its Fintech Reglab*. Consultado en: <https://fintech.adgm.com/adgm-launches-its-fintech-reglab/>



Article 29 Data Protection Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for The Purposes of Regulation* 2016/679, 17/EN. WP 251rev.01, 6 de febrero de 2018.

Banco Mundial, *World Development Report 1997: The State in a Changing World*. Washington: Banco Internacional de Reconstrucción y Fomento / Banco Mundial, agosto de 1997. Consultado en: <http://documentos.bancomundial.org/curated/es/701691468153541519/pdf/173000WDR0SPANISH0Box128708B00PUBLIC0.pdf>

Barrat, J. 2013. *Our Final Invention: Artificial Intelligence and the End of the Human Era*. Oxford, RU: Oxford University Press.

Bartlett, R., A. Morse, R. Stanton y N. Wallace. 2019. *Consumer-Lending Discrimination in the FinTech Era*. Consultado en: <https://faculty.haas.berkeley.edu/morse/research/papers/discrim.pdf>

Bostrom, N. 2014. *Superintelligence: Paths, Dangers and Strategies*. Oxford, RU: Oxford University Press.

Castaño, D. 2019a. Nudge + Código: una arquitectura digital para el precedente judicial. En *XX Jornadas de Derecho Administrativo*, Universidad Externado de Colombia, Bogotá.

-----2019b. *GovTech: Privacidad, legalidad y ética digital*. Alcaldía Mayor de Bogotá.

Cavoukian, A. S.f. *Privacy by Design: The 7 Foundational Principles*. Consultado en: [https://iapp.org/media/pdf/resource\\_center/Privacy%20by%20Design%20%207%20Foundational%20Principles.pdf](https://iapp.org/media/pdf/resource_center/Privacy%20by%20Design%20%207%20Foundational%20Principles.pdf)

Crawford. J. y J. Schultz. 2014. *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 94.

Domingos, P. 2015. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake our World*. Basic Books.

Financial Conduct Authority. 2015. *Regulatory Sandbox*. Financial Conduct Authority, Londres, noviembre. Pub ref. 005147. Consultado en: <https://www.fca.org.uk/publication/research/regulatory-sandbox.pdf>

Financial Conduct Authority. 2016. *Regulatory Sandbox Lessons Learned Report*. Financial Conduct Authority, Londres. Consultado en: <https://www.fca.org.uk/publications/research/regulatory-sandbox-lessons-learned-report>

Floridi, L y J. Cowls. 2019. *A Unified Framework of Five Principles for AI in Society*. 1 Harvard Data Science Review 1, 7.

Floridi, L., J. Cowls, M. Beltrametti et al. 2018. *AI4People - An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*. Minds and Machines, 28: 689 - 707.

Goldberg, K y V. Kumar. 2019. *Cognitive Diversity: AI & The Future of Work*. Tata Communications. Consultado en:  
[https://www.tatacommunications.com/wp-content/uploads/2018/09/Report\\_Cognitive-Diversity\\_AI-and-The-Future-of-Work.pdf](https://www.tatacommunications.com/wp-content/uploads/2018/09/Report_Cognitive-Diversity_AI-and-The-Future-of-Work.pdf)

Goodman, B y S. Flaxman. 2017. *European Union Regulations on Algorithmic Decision-Making and 'a Right to Explanation'*. 38 AI MAG. 50, 55-56.

Hawking, S. 2018. *Brief Answers to the Big Questions*. United Kingdom: John Murray Publishers. Consultado en:  
<https://www.kobo.com/ww/en/ebook/brief-answers-to-the-big-questions>

High-Level Expert Group on Artificial Intelligence (HLEG) of the European Commission on Artificial Intelligence. 2019. *A Definition of Artificial Intelligence: Main Capabilities and Scientific Disciplines*. Consultado en:  
<https://www.aepd.es/sites/default/files/2019-09/ai-definition.pdf>

HM Treasury. 2018. *The Economic Value of Data: Discussion Paper*. Crown.

Information Society Project at Yale Law School & Immuta. 2017. *Governing Machine Learning: Exploring the Intersection Between Machine Learning, Law, and Regulation*. White Paper. Consultado en:  
<https://www.immuta.com/governing-machine-learning-exploring-the-intersection-between-machine-learning-law-and-regulation/>

Jiménez, G. y M. Hagan. 2019. *A Regulatory Sandbox for the Industry of Law*. White Paper. Stanford Law School Legal Design Lab. Consultado en:  
<https://law.stanford.edu/publications/a-regulatory-sandbox-for-the-industry-of-law/>

Jordan, M. I. 2019. *Artificial Intelligence: The Revolution that hasn't Happened Yet*. Harvard Data Science Review, 18. Consultado en:  
<https://hdsr.mitpress.mit.edu/pub/wot7mkc1>

Kaminski, M. E. 2019. *The Right to Explanation, Explained*. 34 Berkeley Tech. L. J. 189.

Katyal, S. 2019. *Private Accountability in the Age of Artificial Intelligence*. 66 UCLA L Rev 54. Consultado en:  
[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3309397##](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3309397##)

Keats-Citron, D. 2008. *Technological Due Process*, 85 Wash. L. Rev. 1249.

Keats-Citron, D y F. Pasquale. 2014. *The Scored Society: Due Process for Automated Predictions*, 89 Wash. L. Rev. 1.

Kroll, J., J. Huey, S. Barocas et al. 2017. *Accountable Algorithms*, 165 U. Pa. L. Rev. 633. Consultado en:  
[https://scholarship.law.upenn.edu/penn\\_law\\_review/vol165/iss3/3/](https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/)

Levine, D.S. *Secrecy and Unaccountability: Trade Secrets in Our Public Infrastructure*, 59 Fla. L. Rev. 135, 139.

McKinsey Global Institute. 2017. *Artificial Intelligence: The Next Digital Frontier?* Discussion Paper.

Mittelstadt, B. D. y L. Floridi. 2016. *The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts*. Sci Eng Ethics, 303.

Monetary Authority of Singapore. 2018. *Regulatory Sandbox*. Consultado en: <https://www.mas.gov.sg/development/fintech/regulatory-sandbox>

OECD (Organisation for Economic Cooperation and Development). 2019a. *Recommendations of the Council on Artificial Intelligence*, OECD/LEGAL/0449. Consultado en: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

-----, 2019b. *Artificial Intelligence in Society*, OECD Publishing, Paris. Consultado en: <https://doi.org/10.1787/eedfee77-en>

-----, 2015. *Data-driven Innovation: Big Data for Growth and Well-being*. OECD Publishing. Consultado en: [https://read.oecd-ilibrary.org/science-and-technology/data-driven-innovation\\_9789264229358-en#page1](https://read.oecd-ilibrary.org/science-and-technology/data-driven-innovation_9789264229358-en#page1)

-----, 2014. 'Data-driven Innovation: Big Data for Growth and Well-being'. OECD Publishing, 10.

O'Neil, C. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Journal of Information Privacy and Security, 13:3, 157-159,

Rubinstein, R., D. Lee y P. M. Schwartz. 2008. *Data Mining and Internet Profiling: Emerging Regulatory and Technological Approaches*. 75 U. Chi. L. Rev. 261.

Russell, S. 2019. *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking

-----, 2017. *Provably Beneficial Artificial Intelligence, The Next Step: Exponential Life*. BBVA, OpenMind. Consultado en: <https://www.bbvaopenmind.com/wp-content/uploads/2017/01/BBVA-OpenMind-Provably-Beneficial-Artificial-Intelligence-Stuart-Russell.pdf>

Russell, S., D. Dewey y M. Tegmark. 2015. *Research Priorities for Robust and Beneficial Artificial Intelligence*. AI Magazine.

Shaw, G. 2019. *The Future Computed: AI & Manufacturing*. Microsoft Corporation.

Singapore Digital, *InfoComm Media Development Authority y Personal Data Protection Commission*. 2020. Model: Artificial Intelligence Governance Framework, 2nd ed., enero 21. Consultado en: <https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf>

Solove, D. J. 2001. *Privacy and Power: Computer Databases and Metaphors for Information Privacy*. 53 Stan. L. Rev. 1393.

Selbst, A. D. y S. Barocas. 2018. *The Intuitive Appeal of Explainable Machines*. 87 Fordham. L. Rev. 1085.

Selbst, A. D. y J. Powles. 2017. *Meaningful Information and the Right to Explanation*, 7 Int'l Data Privacy L. 233, 235.

Stephens-Davidowitz, S. 2017. *Everybody Lies: Big Data, New Data, and What the Internet can Tell us about who We Really Are*. New York: HarperCollins Publishers.

Tene, O. y J. Polonetsky. 2013. *Big Data for All: Privacy and User Control in the Age of Analytics*. 11 Nw.J. Tech.& Intell.Prop.239.

----- . 2012. *Privacy in the Age of Big Data: A Time for Big Decisions*, 64 Stan. L. Rev. Online 63, 65-66.

Wachter, S., B. Mittelstadt y L. Floridi. 2017. *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*. 7 Int'l Data Privacy L.

Watkins, P., E. Daniels y S. Slayton. 2018. *First in the Nation: Arizona's Regulatory Sandbox*. *Stanford Law & Policy Review*, 29 (1): 1-17. Consultado en: <https://law.stanford.edu/publications/first-in-the-nation-arizonas-regulatory-sandbox/>

Zarsky, T. Z. 2011. *Government Data Mining and Its Alternatives*. 116 Penn. St. L. Rev, 285.

**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.