El Derecho de las TIC en Iberoamérica

Obra Colectiva de
F.I.A.D.I.
Federación Iberoamericana de
Asociaciones de Derecho e Informática

Marcelo Bauzá Reilly
Director

LA LEY URUGUAY

Queda hecho el depósito que indica la ley.

Impreso en Uruguay

I.S.B.N. 978-9974-900-17-2

URUGUAY

La FIADI es una Organización No Gubernamental Internacional sin fines de lucro, que emerge en 1984 durante el I Congreso Iberoamericano de Informática Jurídica celebrado en Santo Domingo, República Dominicana. Su nacimiento se produce en el marco de las actividades internacionales que promovía en esa época el Centro Regional para la Enseñanza de la Informática, buscando extender los procedimientos y las técnicas informáticas en la Administración, la Justicia, la Educación, la Cultura y el Trabajo, en favor del bienestar y libertad de los pueblos de la región. Bajo tales premisas, que se mantienen, la FIADI es al día de hoy la más antigua asociación académica regional en este campo, sin parangones considerando su amplitud territorial y su sostenida permanencia en el tiempo.

Actualmente y desde algunos años, la FIADI es una ONG Internacional con personería jurídica reconocida por Resolución Ministerial No. 634 del 15 de diciembre de 2015, del Ministerio de RR.EE. de la República Oriental del Uruguay.

Todos quienes quieran comunicarse con la Federación, solicitar información, postular a una membresía, etc., pueden hacerlo dirigiéndose a las formas de contacto previstas en www.fiadi.org, o a cualquiera de sus Directivos actuales (mandato en ejercicio 2018-2021)

Marcelo Bauzá Reilly – Presidente – presidencia@fiadi.org

Mariliana Rico Carrillo – Secretaria General – secretaria@fiadi.org

Ernesto Ibarra Sánchez – Vicepresidente de Relaciones Institucionales, Recursos Económicos y Transparencia – vp.relacionesinstitucionales@fiadi.org

Federico Bueno de Mata – Vicepresidente Formación y Desarrollo de Capital Humano – vp.formacion@fiadi.org

Bibiana Luz Clara – Vicepresidente de Investigación e Innovación – vp.investigacion@fiadi.org

Índice

Pág.

CONTRIBUCIONES

Capítulo XXXII

Derecho y garantías ante el uso público y privado de inteligencia artificial, robótica y big data (*)

Lorenzo Cotino Hueso

1. Una aproximación a los algoritmos, Inteligencia Artificial (IA) robótica y big data y a sus riesgos e impacto

1.1. Aproximación a conceptos convergentes. IA, sistemas de autoaprendizaje, robots, big data

Un algoritmo es una secuencia de pasos para resolver un problema, comandos para que una computadora transforme un input en output. Uno o más se combinan e integran los programas informáticos desde hace décadas. La Comisión Europea en su Comunicación *IA para Europa* (COM(2018) 237 final, de 25 de abril de 2018 señala que "El término "IA" (IA) se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción –con cierto grado de autonomía– con el fin de alcanzar objetivos

(*) El presente estudio es resultado de investigación del proyecto "Derecho y Big Data", Grupo de Investigación en Derecho Público y TIC como investigador de la Universidad Católica de Colombia. Asimismo, se realiza en el marco del proyecto "La regulación de la transformación digital y la economía colaborativa" PROMETEO/2017/064 Generalitat Valenciana y en el marco de ayuda de la Generalitat Valenciana para la estancias de personal investigador en empresa (AEST/2019/013).

específicos." Tal autonomía se define como "capacidad para realizar tareas previstas en función del estado y la percepción actuales, sin intervención humana" (ap. 2.2). Norma ISO 8373 de 2012 sobre Robots y dispositivos robóticos (ver también Stanford University, 2016: 12-14). Los sistemas IA pueden ser deterministas, de modo que su respuesta queda totalmente predeterminada por los algoritmos, que pueden ser muy complejos. El sistema IA puede ser no determinista, esto es, capaz de dar respuestas diferentes e imprevisibles en razón del aprendizaje y las circunstancias y entorno cambiante.

Se habla de "aprendizaje automático" o *machine learning* cuando los resultados de los algoritmos no dependen de lo que los humanos hayan especificado de antemano. Se suministran datos para que los sistemas y algoritmos identifiquen patrones y correlaciones, aprendan de ellos y generen nuevas relaciones, todo ello para hacer predicciones o recomendaciones. Mientras se ejecutan, los humanos no están controlando y en razón la naturaleza de la "caja negra", los resultados no siempre son intuitivamente explicables. El aprendizaje profundo o *deep learning* tiene aún menor intervención humana y está inspirado en el funcionamiento de redes neuronales de nuestro cerebro. Los grandes datos van pasando por distintas "capas" en la que se aplican reglas de aprendizaje, modelos para que pueda evaluar ejemplos e instrucciones para los resultados se vayan comparando y ajustando en cascada. El sistema aprende y utiliza lo aprendido para muy diversas finalidades.

En las *Normas de Derecho civil sobre robótica. Resolución del Parlamento Europeo*, de 16 de febrero de 2017 (en adelante, resolución sobre robótica) la robótica es la "capacidad de aprender de la experiencia y tomar decisiones cuasi independientes— ha hecho que estos robots se asimilen cada vez más a agentes que interactúan con su entorno y pueden modificarlo de forma significativa". Por su parte, la ya referida ISO 8373 de 2012 define robot como "mecanismo accionado programable en dos o más ejes con un grado de autonomía, que se mueve dentro de su entorno, para realizar las tareas previstas". Finalmente y con carácter convergente, cabe hablar de un internet de las cosas robóticas ("Internet of Robotic Things, IoRT, Simoens)".

Pues bien, todas estas tecnologías convergentes "beben" del big data o de los macrodatos. Se habla de las "V": volumen, variedad, velocidad y valor, a las que se añaden entre otras, la veracidad. El big

data integra asimismo el propio tratamiento masivo de los datos. En la también importante:

Resolución del Parlamento Europeo *de 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley* (en adelante, resolución sobre macrodatos) se afirma:

> "el concepto de macrodato se refiere a la recopilación, análisis y acumulación constante de grandes cantidades de datos, incluidos datos personales, procedentes de diferentes fuentes y objeto de un tratamiento automatizado mediante algoritmos informáticos y avanzadas técnicas de tratamiento de datos, utilizando tanto datos almacenados como datos transmitidos en flujo continuo, con el fin de generar correlaciones, tendencias y patrones (analítica de macrodatos)" (letra A).

1.2. Riesgos, peligros e impactos de estas tecnologías

Nadie duda de que estamos ya o en la antesala de una completa transformación, o más bien, revolución digital, de una cuarta revolución industrial en razón de las variadas tecnologías disruptivas. Las posibilidades y usos de la IA es prácticamente ilimitada y sin duda va a implicar un antes y un después para la humanidad a lo largo del siglo XXI. Quedan afectados todos los órdenes (Stanford University, 2016: 18-41): trabajo, mercado, investigación, educación, salud, asistencia social, medio ambiente, márquetin, eficiencia energética, transporte y ciudades inteligentes, transparencia y lucha contra el fraude, salvamento, reacción a catástrofes, aplicación de la ley, elaboración de políticas y toma de decisiones. En todos los ámbitos la IA pueden mejorar y mucho la innovación, eficiencia, eficacia, competitividad o calidad de las decisiones. Y todo sin entrar ahora en el llamado post o transhumanismo como mejora o aumento del propio ser humano a través de las tecnologías cada vez más cercanas o ya directamente dentro de cuerpo y mente humanas (tecnologías *onlive*).

La IA implica incuestionables ventajas, pero también riesgos, que pueden ser incluso existenciales. Aunque no hay que dejarse llevar en exceso, ha sido objeto de atención especial en la literatura y en el cine el temor a que la IA quede fuera de control humano. Se habla (Kurzweil, Bolstrom, Musk, Hawking, etc.) del momento de la "singularidad tecnológica", de una "IA fuerte" (*strong* AI), la también llamada

"explosión" de la IA, o el "giro traicionero" "treacherous turn" hacia una superinteligencia o inteligencia maligna en términos de Bolstrom, el famoso Director del *Future of Humanity Institute* de la Universidad de Oxford. Desde el Instituto del Futuro de la Vida del MIT (*The Future of Life Institute*) se afirma que "la IA tiene el potencial de volverse más inteligente que cualquier ser humano, no tenemos una manera segura de predecir cómo se comportará." Elon Musk (*Tesla, SpaceX, Paypal*, etc.) en 2014 advirtió que la IA presenta la "mayor amenaza existencial"; Stephen Hawking en sus últimos años insistió en que la IA puede "destruir", "mejorar" o "reemplazar" a la humanidad. El Grupo de Altos expertos en Ética de la IA de la UE califican como "preocupaciones a largo plazo" (AI-HLEG 2018: 12-13), en referencia a la existencia futura de una conciencia artificial, de agentes morales artificiales ("artificial Moral Agent"). El peligro para la humanidad puede darse en especial por la delegación masiva de los humanos de decisiones a estas superinteligencias que, asimismo, pueden interactuar entre ellas. Y los peligros pueden darse incluso aunque no desarrollen la referida conciencia artificial.

El peligro de una escalada en el desarrollo de armas autónomas llevó a la afirmación de los Principios éticos de IA de Asilomar por la comunidad científica en 2015. Petit (2017: 29) califica de "Existernality" a estos riesgos existenciales para la humanidad. Calo (2017: 431) ha criticado estas visiones tan negativas del "apocalipsis de la IA" que pueden generar una percepción social muy negativa para el desarrollo de la IA. Un desproporcionado temor que distrae a los *policymakers* y desarrollos éticos y jurídicos.

El lado negativo de la IA ha generado reflexiones críticas de interés (*Dilemata*, 2017). Me permito destacar ahora la "algocracia" o la tiranía de los algoritmos (Danaher), la "dictadura de los datos" y también la "paralización de la privacidad" ("*Paralyzing privacy*") a que hacen referencia Mayer-Schönberger y Cukier (2013) . También estos autores nos previenen de*l dataismo* "en Dios confiamos, y para todo lo demás hay que aportar datos". En la misma línea, Harari en su famoso *Homo Deus* o en 2016 la Casa Blanca habló del "fundamentalismo de los datos", esto es, "la creencia de que los números no pueden mentir y siempre representan la verdad objetiva" (White House, 2016: 10). Han también señala que "el conocimiento total de datos es un desconocimiento absoluto en el grado cero del espíritu". También señala que la red se

ha convertido en un "enjambre digital" ("Digital swarm") amorfo, una panóptico digital en el que somos explotados, una "sociedad psicopolítica de la transparencia": el sujeto neoliberal se explota a sí mismo y además lo hace de forma voluntaria. Y lo que es peor, con una actitud "acompañada del sentimiento de libertad", típico del feliz usuario de internet.

En esta larga lista de peligros e impactos, O'Neil (2016) ha llegado a hablar de "Weapons of math destruction". La IA genera cajas de resonancia, sesgos de confirmación y otros tantos efectos. Los algoritmos y la IA pese a que aparentemente dan buenos y rentables resultados, en sus selecciones y perfilados para la toma de decisiones generan falsos positivos y, casi peor, falsos negativos, esto es, excluyen erróneamente a quién buscar como cliente, dar un crédito, un trabajo, acceso a estudios, una subvención, un contrato público, a quién inspeccionar, hacer un seguimiento, registrar, detener, dejar cruzar la frontera, condenar, fijar la condena, la IA decide si dar servicios médicos y qué tratamiento en concreto, a quién no dejar subir a un avión, etc. Ello además genera espirales de sesgo, error y discriminación de las que es muy difícil salir tanto al individuo como para los que son afines a él. No en vano, los resultados los que se nutren los sistemas serán cada vez más sesgados, erróneos o discriminatorios.

El Parlamento UE en su resolución sobre macrodatos ha insistido en que:

> "los datos de capacitación a menudo son de una calidad cuestionable y no son neutrales" (Letra B), la "baja calidad" de los datos o los procedimientos "podrían dar lugar a algoritmos sesgados, correlaciones falsas, errores, una subestimación de las repercusiones éticas, sociales y legales, el riesgo de utilización de los datos con fines discriminatorios o fraudulentos y la marginación del papel de los seres humanos en esos procesos, lo que puede traducirse en procedimientos deficientes de toma de decisiones con repercusiones negativas en las vidas y oportunidades de los ciudadanos, en particular los grupos marginalizados, así como generar un impacto negativo en las sociedades y empresas" (Considerando m).

A la larga lista de peligros e impactos, hay que sumar el especialmente el condicionamiento de la autonomía y la libertad del ser humano, la posible deshumanización por la interactuación con IA, el riesgo

para el sistema democrático y el condicionamiento y manipulación por el acceso a cultura e información. Y obviamente el grave impacto en la privacidad.

Se ha dicho que nos acercamos peligrosamente al "*Robo-Rubicon* (Latiff y McCloskey), que ha llegado el momento de decidir si "¿queremos dirigir la tecnología o queremos que ella nos dirija a nosotros?: Programar o ser programados (Rushkoff). Estamos en el "periodo crítico" para el Supervisor Europeo de Protección de Datos (Dictamen 4/2015: 15) antes de la masiva adopción de estas tecnologías.

2. Respuestas de la Ética de la IA, nuevas fórmulas y principios normativos y autoridades regulatorias

2.1. La utilidad para el Derecho de la Ética de la IA y sus principios

La Ética de la IA es el marco propio de reflexión de algunas cuestiones clave que suscita la IA y es fuente que inspira al Derecho (Cotino, 2019 b). Igualmente un marco ético de la IA tiene un carácter especialmente preventivo que evita que el Derecho -esencialmente reactivo- tenga que aplicarse. Incluso puede verse como un sistema de alerta temprana frente a impactos y riesgos. También puede servir para detectar a tiempo líneas o caminos que aunque sean lícitos se consideran socialmente inaceptables. Al mismo tiempo la ética puede orientar por qué tecnologías apostar y subvencionar por ser preferibles. En muchos casos, tanto para los colectivos implicados en la IA, como para la sociedad general, la ética puede generar una adhesión personal y colectiva que facilitar el cumplimiento normativo. Asimismo, las referencias éticas se suelen adaptar bien a los sistemas *softlaw* de códigos, normas éticas, autorregulación y políticas que se adoptan en escenarios híbridos privados, empresariales, organizaciones corporativas y sectores profesionales. En esta línea, comisiones, comités, autoridades o responsables de Ética coadyuvan a la eficacia misma del Derecho.

En los últimos años se ha dado un auténtico desfile de documentos sobre la ética de la IA internacionalmente desde el Instituto de Ingenieros Eléctricos y Electrónicos y su "diseño éticamente alineado en diciembre de 2016; desde 2017 la Unión Internacional de Telecomunicaciones hace cumbres anuales *AI for Good*, al igual que la ACM's (Association for Computing Machinery). La Alianza sobre la IA (Partnership on AI) desde 2018 proclama 7 principios éticos. Además

de los 23 principios de Asilomar concretados en 2017, la Universidad de Montreal proclamó otros 10 en la "Declaración para un Desarrollo Responsable de la IA" (2017). Naciones Unidas y la Unesco están activos con centro UNICRI de IA y Robótica. En el ámbito de protección de datos, desde 2014 la Conferencia Internacional de Autoridades de Protección de Datos (ICDPPC) ha actuado, destacando la Declaración sobre Ética y Protección de Datos en el Sector de la IA (Bruselas, 2018).

Por cuanto a países de referencia, en EEUU la Comisión Federal del Comercio ha generado importantes documentos. El Australian Information Commissioner adoptó los *Australian Privacy Principles de* 2016. En Reino Unido ha sido especialmente relevantes el informe del ICO (autoridad de información y protección de datos de marzo de 2017) así como los trabajos parlamentarios de la Cámara de los Lores en 2018. Singapur ha hecho de la IA toda la estrategia de país (www.aisingapore.org) y en enero de 2019 lanza la necesidad de un "Modelo de Marco de Gobernanza de IA" vinculada a principios.

En Europa, la Asamblea Parlamentaria del Consejo de Europa adoptó la Recomendación 2102 (2017)1, de 28 de abril de y en 2019 hay una Declaración del Comité de Ministros. Del Consejo de Europa destaca en cualquier caso amplia Carta ética europea sobre el uso de la IA en los sistemas judiciales de 2019.

En cualquier caso, la UE se ha situado en el epicentro mundial de la Ética de la IA como marca propia de una "IA Made in Europe". La atención ha sido muy intensa: las referidas Resoluciones del Parlamento Europeo sobre robótica y sobre macrodatos en 2017; informes o directrices del Grupo del Artículo 29 (en adelante G29-UE) en 2014 sobre big data o sobre decisiones automatizadas en 2018; del Supervisor Europeo de Protección de Datos son importantes los Dictámenes 4 y 7 de 2015 y 8/2016. En el marco de la actuación de la Comisión Europea, en abril de 2018 publicó su Comunicación sobre IA y nombró 52 miembros del Grupo de expertos de alto nivel en IA (AI HLEG). Este grupo en noviembre publicó el "Proyecto de Directrices de Ética para una IA Confiable" (AI HLEG, 2018) y tras consulta, la versión ya definitiva de las Directrices se ha publicado en abril de 2019 (AI HLEG, 2019). Asimismo la Comisión lanzó el "Plan coordinado sobre la IA" de 7 de diciembre de 2018 en el que la IA confiable con Ética y Derecho en el diseño es el emblema.

Es posible destilar algunos contenidos básicos de esta ética de la IA a partir de los muchos documentos. La resolución del Parlamento UE sobre robótica "pone de relieve el principio de transparencia" (nº 12) que como se expondrá, es la clave de bóveda de los cuatro principios éticos esenciales, "los principios de beneficencia, no maleficencia, autonomía y justicia, así como en los principios consagrados en la Carta de los Derechos Fundamentales de la Unión Europea [...] así como en otros principios [...] como la no estigmatización, la transparencia, la autonomía, la responsabilidad individual, y la responsabilidad social" (nº 13). *AI4People* (2018) analiza los más de 50 principios proclamados internacionalmente para destilarlos en los cuatro principios esenciales mencionados, que precisamente traen causa del ámbito de la biomedicina desde los años 2001. Así, haz el bien y no hagas daño (principios de beneficencia y de no maleficencia). El principio de "no maleficencia" (no hacer daño), tiene una funcionalidad esencialmente preventiva: no hacer daños físicos, psicológicos, financieros o sociales. Evitar daños accidentales y deliberados. Como bienes en riesgo: todo derecho fundamental, la "privacidad, seguridad y "precaución de capacidad". El principio de justicia viene a suponer que la IA debe desarrollarse para el bien común y el beneficio de la humanidad, mejorar el bienestar individual y colectivo, generar prosperidad, valor y maximizar la riqueza y sostenibilidad. Deben lograrse bienes y servicios comunes a bajo costo y de alta calidad, alfabetización, se han de eliminar discriminaciones históricas, evitar sesgos, estigmatización y discriminación.

Otro principio ético básico es la libertad y autonomía humana frente a la "autonomía" artificial. En aras de ellos se subraya la necesaria "supervisión humana de la IA ("Gobernanza de la autonomía)" (AI-HLEG 2018: 14-15), bajo el principio de más control a mayor autonomía e impacto social. Cuanto mayor es el grado de autonomía que se otorga a un sistema IA, se requieren más pruebas y auditoría y una gobernanza más estricta. Las tecnologías deben respetar la capacidad humana para elegir y frente a la manipulación, se apuesta por el complejo concepto de *nudging* introducido por Thaler y Sunstein, como auto-empujón, refuerzo positivo indirecto para influir en el comportamiento.

El principio de transparencia es el quinto principio esencial. Ha pasado a considerarse "la pieza crucial que falta en el rompecabezas" pues "complementa los otros cuatro principios" (AI-HLEG, 2018: 10). El mismo incluye no sólo el acceso a funcionamiento y resultados, sino

también la explicabilidad, auditabilidad y trazabilidad. de la ética de la IA (principios de beneficencia, no maleficencia, autonomía y justicia). Más adelante se concreta su alcance.

Además, se han formulado otros principios básicos como precaución; participación; rendición de cuentas; seguridad; reversibilidad e inversión de acciones y volver a la fase «buena» de su trabajo; privacidad y maximización de beneficios y reducir al mínimo los daños (Anexo resolución Parlamento UE sobre robótica). El grupo de altos expertos de la UE (AI-HLEG 2018: 14-18) entre otros principios afirma la responsabilidad (1º); gobernanza de los datos y anonimización, evitación de datos maliciosos (2º)y el diseño para todos (3º). Igualmente el principio (8º) de "robustez" ("Robustness") que incluye la fiabilidad y reproducibilidad ("reliability & reproducibility"), la exactitud de los datos y sistemas. También, se exige un "Plan de retroceso" ("Fall back plan").

2.2. Nuevas fórmulas y principios normativos

Ante la IA que va generando retos y riesgos tan cambiantes, es bien posible que haya que introducir nuevas formas y mecanismos de regulación y gobernanza. Una vida, sociedad, modernidad e incluso maldad líquida (Bauman) en la que lo único permanente es el cambio constante, exige un Derecho líquido o biodegradable, experimental en beta continuo. Más allá del Derecho tradicional es necesario acudir a "políticas de IA" ("AI policies", Calo, 2017, 407) así como a nuevas fórmulas regulatorias experimentales que incorporen la evaluación y adecuación normativa; fórmulas que faciliten que se den sucesivas versiones y actualizaciones normativas, que remitan a los órganos capacitados técnicamente para la concreción normativa. Se habla de "*regulatory sandboxes"* en referencia a las zonas de arena acotadas para que los niños jueguen. Una *smart regulation* en términos de Zetzsche. Las nuevas formas regulatorias precisas para la IA llevan también a un Derecho nebuloso y un Derecho en red, con alta participación del sector tecnológico, de informáticos y de especialistas, con un peso creciente de estandarización técnica internacional de naturaleza privada, con formas de regulación suaves y modelos de co-regulación trasnacional.

Obviamente hay que estar alerta para no dejar en manos de programadores o directamente de algoritmos las propuestas regulatorias, hay que vigilar los problemas serios de *traducción* de lo jurídico por

los técnicos (Citron, 2007: 1254) y la falta de legitimación democrática tanto de técnicos como del sector privado. También hay que prever las capturas regulatorias o el capitalismo clientelar. Y estas influencias pueden venir tanto desde el sector de la IA pero también en sentido contrario desde el sector social o empresarial que queda desplazado por las tecnologías emergentes. También los crecientes códigos corporativos y éticos en sectores tecnológicos pueden ser contrarios a la competencia.

En la regulación de la IA también es necesario adoptar técnicas relativas al principio cautela o precaución ya proclamado por la Comisión Europea en su Comunicación *COM/2000/0001* de 2000. Hay que identificar riesgos, quienes los aleguen deben demostrarlos y, como consecuencia, se puede exigir que el productor, el fabricante o el importador demuestren la ausencia de peligro. Es exigible la participación de los implicados y transparencia en los procesos e investigación. Las decisiones deben ser proporcionales, no discriminatorias y coherentes con otras decisiones.

Igualmente es esencial tener presente el genérico principio de responsabilidad y evitar riesgos irreversibles. Así, "Las generaciones actuales tienen la responsabilidad de garantizar la plena salvaguardia de las necesidades y los intereses de las generaciones presentes y futuras." (art. 1 Declaración sobre las responsabilidades de las generaciones actuales para con las generaciones futuras, UNESCO 1997) y "deben esforzarse por asegurar el mantenimiento y la perpetuación de la humanidad, respetando debidamente la dignidad de la persona humana." (art. 3).

Y especialmente, ante la IA hay que asumir técnicas de responsabilidad "proactiva, "demostrada" o "accountability". A modo de las leyes de la robótica de Asimov, se trata de la inserción del cumplimiento ético y legal en el mismo "Código", en los conocidos términos de Lessig. Se trata del modelo ya impulsado por el Reglamento (UE) 2016/679, de 27 de abril de 2016, europeo de protección de datos relativo (en adelante RGPD-UE), que apuesta por la protección de datos desde el diseño y por defecto (art. 25) o la evaluación de impacto de protección de datos (artículo 35). Precisamente el "*Ethics & Rule of law by design* X-by design" (AI-HLEG 2018: 19) es la enseña de la "IA confiable" "Made in Europe" (AI-HLEG 2018: 21, 29) que está en la base del Plan coordinado sobre la IA de diciembre de 2018 de la Comisión Europea). Lejos de proclamaciones generales, las Directrices de Ética para una IA Confiable

adoptadas en abril de 2019 (AI-HLEG 2019: 26-31), incluyen un *check list - trustworthy AI assessment list* con unas 150 cuestiones a formularse en cualquier proyecto de IA. Aunque no se trate de una norma jurídica, en buena medida va a pasar a exigirse en futuros proyectos de IA que se lleve a cabo el análisis de impacto ético. Y es que este listado tiene claro paralelismo con las listas previstas para los estudios de impacto de protección de datos de la UE (art. 35. 4º RGPD-UE).

2.3. Autoridades sectoriales y especializadas para la IA

Un elemento importante de la "regulación", adopción de políticas o gobernanza en materia de IA, big data y robótica es la determinación del *regulador*, esto es, organizaciones sectoriales y de autoridades con competencias que han de fijar criterios, controlar y supervisar el sector. Las posibilidades son muchas. El Parlamento británico habló de la necesidad de una autoridad robótica del tipo de la Civil Aviation Authority; la Cámara de los Lores constituyó en 2017 el "Artificial Intelligence Committee" y en 2018 se apuesta por que Centro de Ética e Innovación de los Datos de Reino Unido elabore un código IA para el sector público y privado. La Comisión Europea lanzó una "Alianza europea de la IA" para integrar el sector. Y diversos grupos de expertos tienen importancia: el mencionado Grupo de Expertos de Alto Nivel sobre IA; el European Group on Ethics in Science and New Technologies (EGE). Para EEUU, se ha afirmado la conveniencia de a modo de una FDA (Food and Drug Administration). Se apuestas por una uniformidad federal con juicio experto, independencia y conocimiento del mercado, un modelo que no frene la innovación. Esta autoridad habría de contar con tres poderes: el de organizar y clasificar algoritmos en categorías regulatorias por su diseño, complejidad y potencial de daño (tanto en uso ordinario como a través del uso indebido); la capacidad de prevenir y una amplia autoridad para imponer requisitos de divulgación y restricciones. En cualquier caso, en EEUU es bien posible que proyecte a la IA su ámbito de actuación la Comisión Federal de Comercio (FTC), de tradicional actividad en los ámbitos de privacidad y competencia, o la Comisión Federal de Comunicaciones (FCC).

Considero que en el futuro próximo es muy posible que se redefinan y vayan nutriéndose de competencias en materia de IA y big data las ya existentes en diversas áreas: competencia, comunicaciones, transparencia, fraude, fiscales, etc. Y en todo caso, han de jugar un muy

importante papel las autoridades independientes cercanas al sector de las nuevas tecnologías, en particular las autoridades de protección de datos y transparencia.

Las autoridades responsables habrán de contar con fuertes facultades de acceso y conocimiento de los algoritmos, lógicas y los grandes datos que manejan tanto el sector privado como el público. Se trata de una vía importante para conciliar la necesidad de control y supervisión del cumplimiento normativo con las obligaciones de confidencialidad, secreto, derechos de propiedad, propiedad industrial, intelectual legítimos del sector respecto de sus datos, algoritmos e IA. Asimismo, va a ser muy importante dotar a estas agencias de personal muy cualificado a estas instituciones, así como prevenir la huida de los mismos a los sectores que tienen que controlar, así como tomar precauciones contra fenómenos como las *puertas giratorias*.

3. La aplicación a la IA del régimen de protección de datos y las garantías añadidas del nuevo "derecho" a no ser sometido a decisiones automatizadas

3.1. La aplicación del régimen general protección de datos para la IA

La IA atrae casi por defecto la aplicación del régimen de la protección de datos. Y en ocasiones, es casi el único régimen jurídico hoy día claramente aplicable. En muchos casos la IA implica la elaboración de perfiles, esto es, evaluación automatizada de personas por cuanto su rendimiento profesional, situación económica, salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos de dicha persona física, etc. (art. 4. 4º RGPD-UE). Asimismo, la IA supone también en muchos casos decisiones automatizadas basadas en datos directos, observados o inferidos de las personas (G29-UE, 2018: 7-8). Para que sea aplicable el régimen de protección de datos debe darse la premisa de que los variados macrodatos que *alimentan* la IA sean datos de personas identificadas o identificables, o reidentificables. No se aplicará la normativa si se da una anonimización que garantice que los datos no vuelvan a ser personales (AEPD-ISMS, 2017, pp. 40 y ss.; Stalla-Bourdillon y Knight, 2017). A este respecto hay que seguir especialmente el Dictamen 5/2014, de 10 de abril, del G29-UE sobre anonimización.

Si hay tratamiento de datos, no es difícil que el RGPD-UE extienda mundialmente su aplicación en muchos casos de uso de la IA. No en vano el RGPD-UE rige respecto de los tratamientos relativos al control del comportamiento de personas en la UE y también se extiende a tratamientos respecto de toda oferta de bienes o servicios dirigidos a la Unión Europea (art. 3. 2º).

Si se aplica la normativa de protección de datos a la IA hay que partir del cumplimiento de los principios, la legitimación del tratamiento, los derechos, la responsabilidad proactiva y privacidad en el diseño o el régimen de las transferencias internacionales de datos. Asimismo y por defecto se exigirá el estudio de impacto. No es posible abordar exhaustivamente estas cuestiones aunque existen importantes referentes a seguir (G29-UE, 2018; AEPD-ISMS 2017).

Así, el tratamiento de datos que implique la IA ha de ser legítimo, como puede ser a través del consentimiento y el contrato. Sin embargo, en el ecosistema del big data, la IA y el aprendizaje autónomo es muy difícil consentir –o informar- respecto de unas finalidades que por lo general ni se conocen, ni se sospechan. Tampoco es fácil informar sobre la localización, transferencias internacionales o tiempo del tratamiento, etc. El artículo 7. 4º RGPD-UE puede generar incertidumbre sobre el consentimiento. Cabe recordar que no hay consentimiento libre si se recaban datos "que no son necesarios" para la ejecución de un contrato. Y no olvidemos que muchísimos datos personales de los que se *alimenta* la IA se recaban masivamente sin relación con el servicio que se presta al usuario que los facilita (por ejemplo, el acopio de ingentes datos en las redes sociales, juegos online, servicios de vídeo y otros contenidos y un largo etcétera).

Más allá del consentimiento, se han de hacer esfuerzos para legitimar el tratamiento de datos a favor de las muchas ventajas que la IA puede suponer para la sociedad y los legítimos intereses públicos y privados en juego. Y la legitimación de grandes tratamientos de datos es necesaria incluso en el ámbito de los datos especialmente protegidos, como los datos de salud o biométricos. El difícil equilibrio entre los derechos, intereses y bienes en juego pasa en buena medida porque la legitimación del tratamiento *sea a cambio* de garantías o medidas compensatorias adecuadas. Luego se apunta un abanico de posibles garantías que pueden facilitar la legitimación.

El legislador puede legitimar la IA con ponderación y garantías (art. 6. 3º RGPD-UE). Ahora bien, en España la reciente Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos brinda dos ejemplos, una de cal y otra de arena. La positiva cal: la Disposición adicional 17ª facilita y legitima la investigación en materia de salud y biomédica sin consentimiento, una investigación que en muy buena medida puede implicar big data e IA. Y lo hace detallando no pocas garantías y concreciones respecto de la información y transparencia obligatorias, la exigencia de actuación de comités, la importante técnica de la seudoanimización y garantías frente a la reidentificación,, garantías de confidencialidad, de seguridad, de derechos, evaluación de impacto, designación de representante legal, etc. Por el lado negativo, la negativa *arena*: en mayo de 2019 el TC español ha considerado inconstitucional la misma Ley Orgánica 3/2018 por cuanto permitía a los partidos políticos el perfilado y tratamiento masivo de datos ideológicos de la población. A diferencia del caso de salud, simplemente señalaba que se haría "con las garantías adecuadas", pero sin detallarlas.

Otra vía de legitimación del tratamiento de datos que supone la IA es la del "interés legítimo". Que se acepte esta legitimación por el regulador exige que se den *a cambio* toda una serie de garantías (art. 6. 1º f) RGPD-UE). Más garantías según sea más detallado y exhaustivo el nivel de detalle del perfilado de datos, o el mayor impacto y consecuencias que sobre las personas tenga el perfilado o la decisión automatizadas (G29-UE, 2018:15). También mayores garantías según la peligrosidad por la procedencia o la naturaleza de los datos empleados. En el caso de menores o de mercadotecnia o publicidad resultará más difícil que el uso de IA se considere de interés legítimo. Asimismo puede incluso haber prohibiciones concretas o excepcionales garantías para los usos de IA a partir de datos especialmente protegidos.

También cuando la IA está sujeta a la protección de datos deben implementarse los derechos de acceso, rectificación, supresión, olvido, oposición o nuevos derechos reconocidos como la portabilidad, entre otros (art. 20 RGPD-UE, o en Brasil art. 16 Ley nº 13.709, de 14 de agosto de 2018). Y en modo alguno es fácil hacer efectivos tales derechos en el ámbito de los macrodatos y la IA (G29-UE, 2018:17-21).

En todo caso, el cumplimiento del régimen de protección de datos no puede hacerse depender ni del consentimiento ni del ejercicio de derechos y garantías por los interesados. Los individuos por lo general

son ignorantes, inconscientes e incluso indolentes ante la IA. Según se ha señalado, el modelo a seguir y reforzar en el ámbito de la IA es la exigencia del cumplimiento proactivo de las obligaciones legales preventivas de privacidad en el diseño y por defecto y el logro de la minimización del tratamiento de datos. Asimismo el uso de la IA respecto de humanos obliga efectuar la evaluación de impacto de protección de datos (art. 35 RGPD-UE), esto es, el análisis y descripción de todas las operaciones, su necesidad y la proporcionalidad y la evaluación de los riesgos (al respecto, G29-UE, 2018: 3 y ss. De igual modo, en el ámbito de la IA hay que subrayar las exigencias de seguridad, fomentarse el cifrado, técnicas de seudoanonimización o extremar las garantías para evitar que se permitan o faciliten las "puertas traseras" (resolución Parlamento UE sobre macrodatos UE y AEPD-ISMS 2017).

3.2. Las garantías añadidas del nuevo "derecho" a no ser sometido a decisiones automatizadas

El uso de la IA respecto de los humanos es el ámbito potencial de proyección del "derecho" a no ser sometido a decisiones automatizadas. Se trata de un derecho reconocido en el artículo 22 RGPD) de la UE: De especial referencia al respecto son las "Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679" del ya extinto Grupo del Artículo 29, de 3 de octubre de 2017, revisadas el 6 de febrero de 2018.

Asimismo el artículo 11 de la Directiva (UE) 2016/680 para el ámbito judicial, policial y de seguridad. En mayo de 2018 este derecho se ha extendido internacionalmente con la muy reciente modificación del Convenio 108 del Consejo de Europa que lo reconoce, también como "derecho", en su nuevo artículo 9. 1º, al igual que los Estándares de Protección de Datos para los Estados Iberoamericanos de 2017 (nº 29).

En concreto, a cada individuo se le reconoce el "derecho" a " no estar sujeto a una decisión que le afecte significativamente basándose únicamente en un procesamiento automatizado de datos sin que se tengan en cuenta sus opiniones"

"1. Every individual shall have a right:

a. not to be subject to a decision significantly affecting him or her based solely on an automated processing of data without having his or her views taken into consideration...".

Vid Consejo de Europa - No. [223] Protocolo por el que se modifica el Convenio para la Protección de las Personas con respecto al procesamiento automático de datos personales [Estrasburgo, 10.X.2018] https://www.coe.int/en/web/conventions/new-treaties

Asimismo, también está reconocido como derecho en el artículo 20 de la reciente Lei nº 13.709, de 14 de agosto de 2018 de Brasil. Con la IA adquiere protagonismo un derecho que guardaba telarañas sin tener relevancia alguna en España desde 1992, pues ya estaba regulado en el artículo 12 de la Ley Orgánica 5/1992, de 29 de octubre, luego en el artículo 13 de la ya derogada Ley Orgánica 15/1999, de 13 de diciembre.

Para una mejor comprensión jurídica de este derecho, puede ser útil partir de que, en muy buena medida, se afirma como derecho subjetivo lo que es un conjunto de especialidades y garantías, las cuales se superponen a las que garantiza el más general derecho fundamental de protección de datos. Este conjunto de garantías específicas de este "derecho" entroncan con el general derecho de protección de datos por cuanto se "refuerza la idea de que sea el interesado quien tenga el control sobre sus datos personales" (G29-UE (2018: 22). No obstante, pese al literal reconocimiento como "derecho", para el G 29 (2018: 29) el artículo 22 RGPD-UE no "deba interpretarse como un derecho".

En cualquier caso, la clara intención de este "derecho" es que las decisiones automatizadas relevantes, por su sensibilidad o particularidad, tienen ser compensadas con garantías especiales.

Como punto de partida, hay que delimitar cuándo opera este nuevo derecho. El mismo es relativo a "una decisión basada *únicamente* en el tratamiento automatizado, incluida la elaboración de perfiles, que *produzca efectos jurídicos* en él *o le afecte significativamente* de modo similar". Se consideran ejemplos típicos "la denegación automática de una solicitud de crédito en línea" o "los servicios de contratación en red en los que no medie intervención humana alguna" (considerando 71 RGPD-UE). El G29-UE (2018: 24) incluye como "significativas" decisiones automatizadas de crédito, servicios sanitarios, oportunidades laborales o de acceso a la educación. En el artículo 20 de la ley brasileña quedan "incluidas las decisiones destinadas a definir su perfil personal, profesional, de consumo y de crédito o los aspectos de su personalidad".

Del lado contrario, quedarían fuera de las garantías de este particular derecho las decisiones automatizadas que no se consideren "significativas" o relevantes. Es dudoso que hoy día este derecho alcance millonarias y rutinarias decisiones automatizadas de perfilado e individualización masivo de información, contenidos y servicios. Estas decisiones pueden determinar la percepción, conocimiento y sentimientos de cada persona. Asimismo y en conjunto, pueden condicionar el sistema político o el mercado mismo. En todo caso, el G29-UE (2018: 24-25) considera que el perfilado en la publicidad en línea sí que puede ser "significativo" y por tanto, este derecho sería aplicable en razón del "intrusismo" en el perfilado, las expectativas y deseos de las personas afectadas; la forma en que se presenta el anuncio; o el uso de conocimientos sobre las vulnerabilidades de los interesados." También este derecho se aplicaría por considerarse "significativas" las "decisiones automatizadas que provoquen diferencias de precios [...] prohibitivamente elevados".

Este derecho es aplicable respecto de las decisiones "*únicamente*" automatizadas. Así pues, quedarían fuera del ámbito de este derecho las decisiones basadas en IA pero en las que se da alguna intervención humana, aunque sean decisiones muy relevantes o produzcan efectos jurídicos en las personas. Sí que quedarían fuera de este derecho las decisiones si "un ser humano revisa y tiene en cuenta otros factores para tomar la decisión final". No obstante, y esto es importante, "si alguien aplica de forma rutinaria perfiles generados automáticamente a personas sin que ello [la revisión humana] tenga influencia real alguna en el resultado, esto seguiría siendo una decisión basada únicamente en el tratamiento automatizado". Para que no rija este derecho, la intervención humana ha de ser "significativa, en vez de ser únicamente un gesto simbólico" y llevada a cabo por "persona autorizada y competente". Hay que señalar que el estudio de impacto que debe hacerse ha de registrar el grado de intervención humana. (G29-UE, 2018: 23).

Pese a que el derecho en principio no alcance a las decisiones parcialmente automatizadas, el legislador puede aplicar garantías en algunos supuestos. Así, en Francia, la ley de la República Digital, ley n° 2016-1321 de 7 de octubre de 2016 prohíbe las decisiones semiautomatizadas en el ámbito de la justicia o incluye especiales limitaciones para el ámbito administrativo (artículo 10.1º y 3º respectivamente de la Loi n° 78-17 de 6 de enero de 1978).

Una vez analizados los presupuestos de aplicación de este derecho, cabe delimitar su contenido y facultades. Así, como punto de partida, este derecho implica -y sin necesidad de que lo solicite el interesado- una prohibición general de las decisiones basadas únicamente en el tratamiento automatizado. Esto se traduce en que se exigen algunos requisitos especiales para la legitimación de tratamientos sólo automatizados, como pueda ser un contrato o el consentimiento explícito. Si se da esta legitimación especial, además, debe reconocerse el "derecho a obtener intervención humana por parte del responsable, a expresar su punto de vista y a impugnar la decisión" art.22. 3º RGPD-UE). Para el G29-UE (2018: 39) sólo éste concreto derecho subjetivo es un "derecho activo".

La prohibición general de decisiones sólo automatizadas también se levanta si hay una legislación que, además de legitimar el tratamiento, incluya garantías compensatorias.

Debe señalarse que este "derecho" también supone una prohibición más intensa de tratamientos automatizados si se basan en datos especialmente protegidos. No obstante, el consentimiento explícito o una legislación específica en razón de un "interés público fundamental" pueden levantar esta prohibición (art. 9. 2 a) y g) RGPD-UE) En los dos casos serán más intensa la obligación de que el tratamiento automatizado sensible se compense con especiales garantías. El considerando 71 del RGPD-UE afirma que una decisión sólo automatizada relevante "no debe afectar a un menor"; sin embargo esta prohibición no está en el texto del RGPD-UE. Así pues, no hay una prohibición absoluta pero sin duda, también habrá que aumentar las garantías compensatorias en estos supuestos.

Pues bien, cabe recordar que el G29-UE (2018: 37-38) detalla diversas de estas garantías. Se trata de una serie de medidas esencialmente "para garantizar que las personas reciben un trato justo y no discriminatorio", sin "resultados discriminatorios, erróneos o injustificados". Para ello, se señalan como "buenas prácticas" garantías como:

- controles periódicos de calidad de sus sistemas.
- Auditorías algorítmicas: comprobación del funcionamiento y resultados de los algoritmos utilizados y desarrollados por los sistemas de aprendizaje automático.

- Si hay elevado impacto sobre las personas, auditorías independientes de terceros, quienes deben acceder a toda la información necesaria del sistema IA.
- Garantías contractuales respecto de los algoritmos de terceros que garanticen comprobaciones y cumplimiento normativo.
- Medidas específicas para la minimización de datos, claros periodos de conservación.
- Técnicas de anonimización y seudoanonimización respecto de la elaboración de perfiles;
- Formas de permitir al interesado expresar su punto de vista e impugnar la decisión
- mecanismo para la intervención humana en determinados casos, como enlaces a recurso o impugnaciones, plazos determinados, contacto para cualquier consulta.

Asimismo se apuntan otras opciones como mecanismos de certificación para operaciones de tratamiento; códigos de conducta para procesos de auditoría; comités de ética para evaluar los daños y beneficios.

Resulta de interés tener en cuenta estas garantías enunciadas por el G29-UE y en paralelo el listado de estudio de impacto ético con 150 preguntas que ha incluido la UE en abril de 2019 (AI-HLEG (2019).

Este derecho respecto de las decisiones sólo automatizadas relevantes implica también particulares deberes de transparencia. Así, se da la concreta obligación de facilitar "información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado" (art. 13. 2º f y 14. 2º g) RGPD-UE). De igual modo, el derecho de acceso permite solicitar dicha información (artículo 15. 1º. h).

El G29-UE (2018: 35) detalla que la información que debe facilitarse es sobre:

- las categorías de datos que se han utilizado o se utilizarán en la elaboración de perfiles o el proceso de toma de decisiones;
- por qué estas categorías se consideran pertinentes;

- cómo se elaboran los perfiles utilizados en el proceso de decisiones automatizadas, incluidas las estadísticas utilizadas en el análisis;
- por qué este perfil es pertinente para el proceso de decisiones automatizadas; y cómo se utiliza para una decisión relativa al interesado.

También se recomienda informar en general respecto de toda decisión automatizada, aunque no sean las protegidas por el artículo 22 y este derecho G29-UE (2018: 27).

Se acaban de enunciar las garantías específicas que implica este "derecho" cuando se dan sus presupuestos. Ahora bien, cabe reiterar que estas garantías en todo caso se añaden o superponen a las que ya confiere el régimen general de protección de datos. Así, respecto de decisiones que no sean sólo automatizadas o no sean relevantes para el afectado, *siempre quedará el París* del régimen general de protección de datos aplicable.

Estas obligaciones de transparencia cabe tenerlas en cuenta con otras obligaciones de transparencia que se analizan *infra*.

4. La no discriminación algorítmica y su minimización

4.1. Errores, sesgos y discriminaciones masivas y no inteligentes

La IA y en particular las tecnologías *onlive* pueden llevar al posthumanismo o transhumanismo, generando una raza superior o mejorada de seres humanos. Y es muy posible que estas mejoras queden limitadas a algunos países y, en especial, a algunos colectivos, generándose una grave discriminación. No obstante, ello trasciende el presente estudio.

Al mencionar los riesgos e impactos de la IA ya se ha hecho referencia a los errores, sesgos y discriminación masivos. Los mismos pueden darse sin intención alguna, incluso con la pretensión de lograr toda la objetividad posible. Ello por lo general se deberá a malas elecciones de datos, datos sesgados contra un grupo, datos pobres, incompletos, incorrectos, desactualizados, mal recopilados, de mala calidad, *vid* FRA. (2018). *#BigData… op. cit.* pp. 4-5. Las mismas series históricas de datos que alimentan al sistema de IA pueden ser reflejo de una realidad social discriminatoria. Lo peor, además, es que se pueden

generar espirales de sesgo, error y discriminación. Los sistemas muy posiblemente acentuarán sus decisiones al nutrirse de nuevos datos cada vez más negativos para los sectores perjudicados. Errores, sesgos y discriminaciones también pueden ser causa de un erróneo diseño y elección de algoritmos, del peso que atribuyen a unos u otros factores o a los errores en el desarrollo de los sistemas de aprendizaje automático. La detección de las causas de sesgos y discriminaciones puede ser realmente compleja. Se precisará acceder en lo posible al sistema y sus registros, así como especialmente analizar los resultados del uso de la IA.

Más allá de errores, el sistema de IA y algoritmos puede estar diseñado para tener en cuenta circunstancias especialmente prohibidas (sexo, raza, religión, salud, etc.). En muchos casos, supondrá un tratamiento de datos especialmente protegidos (art. 9 RGPD-UE), lo cual implica particulares restricciones y garantías y posibles prohibiciones. Al régimen de datos especialmente protegidos se podrán superponer las particulares garantías de las decisiones sólo automatizadas significativas (art. 22. 3º RGPD-UE). Además, los tratamientos diferentes basados en circunstancias especialmente prohibidas son especialmente sospechosos de discriminación. Se presume que no son admisibles y quedan sometidos a un escrutinio o test de admisibilidad más riguroso.

El artículo 11. 3º de la Directiva (UE) 2016/680 sobre tratamientos automatizados en materia penal y de justicia expresamente prohíbe "La elaboración de perfiles que dé lugar a una discriminación de las personas físicas basándose en las categorías especiales de datos personales establecidas en el artículo". Debe señalarse que discriminar está prohibido en cualquier caso. Este precepto puede entenderse como una limitación más intensa si cabe al perfilado de datos en el más sensible ámbito policial y penal, que sólo muy excepcionalmente podrá admitirse. En EEUU la clave jurídica para admitir que los algoritmos y el big data incluyan información racial reside en que el algoritmo no haya elegido el factor raza con finalidad discriminatoria ("because of" "discriminatory purpose"), que es lo que prohíbe la doctrina *McCleskey v. Kemp* 481 U.S. 279, 291-92 (1987). Por el contrario, se podría de admitir que se tome en cuenta el factor raza por ser un elemento que aumenta la precisión "a pesar de" ("in spite of") sus efectos sobre un grupo identificable. Habrá que analizar rigurosamente cada caso concreto el funcionamiento y resultados de la IA, eso sí, bajo la presunción de discriminación.

Hay que advertir que puede que se intente eludir las prohibiciones o limitaciones de utilizar datos especialmente protegidos en los sistemas IA utilizando *proxies* o datos afines. Así, datos de comportamiento, preferencias, geográficos o similares, pueden fácilmente revelar sexo, raza, religión, orientación sexual, etc. En este sentido, habrá que analizar posibles "enmascaramientos" y la elección intencional de factores que están cerca de los prohibidos.

4.2. Minimización de datos y estudio de impacto de discriminación

La resolución del Parlamento UE sobre macrodatos "insta" a "minimizar la discriminación y el sesgo algorítmicos" (nº 20) y afirma también la necesaria "mitigación algorítmica" (nº 21, ver también 32). Subraya especialmente que se incluyan mecanismos de transparencia y rendición de cuentas y la posibilidad de corrección de datos y de recurrir decisiones algorítmicas. Y también los propios algoritmos pueden ser el antídoto contra el sesgo algorítmico y se pueden programar para ignorar o minimizar la importancia que los factores prohibidos en sus decisiones.

La minimización, no obstante, debe garantizar que el sistema inteligente no pierda su eficacia, su propia naturaleza o lleve a generar resultados absurdos. Y es que la eliminación o minimización de algunos datos o factores pueden hacer disfuncionales o incluso inservibles los perfilados, clasificaciones, correlaciones o predicciones.

Jurídicamente, siguiendo técnicas del Derecho antidiscriminatorio, las medidas de corrección de posibles sesgos han de estar bien justificadas en su necesidad, así como en su razonabilidad y proporcionalidad. De lo contrario pueden a su vez constituir un tratamiento discriminatorio. En Estados Unidos, por ejemplo, la sentencia Ricci v. DeStefano, 557 U.S. 557 (2009) 585 señala que para minimizar una posible discriminación debe exigirse una "fuerte base probatoria de que si no se adoptan las medidas se generará una discriminación".

Frente a la discriminación algorítmica cabe proyectar técnicas de "discrimination impact assessments" y de evaluación de impacto de género bien conocidas en la UE. En todo caso, hay que extender el modelo de la responsabilidad proactiva en protección de datos, la no discriminación en el diseño y por defecto, así como medidas concre-

tas en los estudios de impacto. Precisamente, varias de las "buenas prácticas" arriba enunciadas que el G29-UE (2018: 36) incluye como garantías respecto de las decisiones automatizadas son frente a la discriminación. Igualmente, no pocas de las 150 cuestiones del *checklist* en las Directrices para la ética en el diseño en la UE lo son para lograr la exactitud y fiabilidad, integridad, calidad de los datos y para evitar el sesgo y la discriminación (AI-HLEG 2019: 26-31). Así, en el bloque sobre robustez técnica y seguridad (2) se incluyen cuestiones para asegurar la exactitud y fiabilidad (¿Ha implementado medidas para garantizar que los datos utilizados sean completos y estén actualizados ?; ¿Se implementaron medidas para evaluar si se necesitan datos adicionales, por ejemplo, para mejorar la precisión o para eliminar el sesgo ?; ¿Se verificó qué daño se causaría si el sistema de AI hace predicciones inexactas ? ¿Se dispusieron formas de medir si el sistema está haciendo una cantidad inaceptable de predicciones inexactas ? ¿pasos para aumentar la precisión del sistema ?. En el bloque 3 sobre privacidad hay cuestiones sobre la calidad e integridad de los datos (¿Estableció mecanismos de supervisión para la recolección, el almacenamiento, el procesamiento y el uso de los datos ?, ¿Evaluó la medida en que tiene el control de la calidad de las fuentes de datos externas utilizadas ? ¿Ha implementado procesos para garantizar la calidad e integridad de sus datos ? ¿Consideró otros procesos ? ¿Cómo está verificando que sus conjuntos de datos no hayan sido comprometidos o pirateados ?).

Y precisamente hay un bloque sobre "Diversidad, no discriminación y equidad" (5) con 19 cuestiones. Diversas de las cuales giran sobre "Evitación del sesgo injusto" (¿Estableció una estrategia o un conjunto de procedimientos para evitar crear o reforzar un sesgo injusto en el sistema AI, tanto en el uso de los datos de entrada como en el diseño del algoritmo ?, diversidad y representatividad en los datos; inclusión poblaciones específicas, uso de pruebas en las fases de desarrollo, incorporación de mecanismos específicos para que otros puedan identificar problemas de sesgo o discriminación; mecanismos para consultar estos problemas; consideración de afectados indirectos. Evaluación si variación de resultados en mismas condiciones, cómo medirlo. ¿Cómo se ha medido la imparcialidad?

De igual modo se incluyen en este bloque cuestiones sobre accesibilidad y diseño universal (si el sistema de AI es utilizable por personas con necesidades especiales; si la información sobre el sistema de AI es

accesible; si en la elaboración se involucró o consultó a esta comunidad; si el equipo de desarrolladores es representativo de su público objetivo; si se evaluó si podría haber personas o grupos que pudieran verse afectados de manera desproporcionada por implicaciones negativas, si se recibieron comentarios de otros equipos o grupos). El bloque sobre bienestar social y ambiental (6) cuenta con alguna cuestión sobre si, por ejemplo, se tuvo en cuenta el impacto social general, como pérdidas de empleo. Finalmente, en el bloque de responsabilidad (7, auditabilidad y documentación de las compensaciones) también se incluyen cuestiones sobre auditoría independiente, evaluación de riesgo o impacto, capacitación y educación, introducción de mecanismos de "junta de revisión ética de AI" o similares. Vinculado a la minimización del sesgo se hace referencia a la posibilidad de compensaciones "trade-off": ¿Estableció un mecanismo para identificar los intereses y valores relevantes implicados por el sistema de AI y las posibles compensaciones entre ellos?" Asimismo se señala la necesidad de documentar la decisión de compensación.

El artículo 20. 2º de la ley brasileña prevé que "la autoridad nacional podrá realizar auditoría para verificación de aspectos discriminatorios en tratamiento automatizado de datos personales." (art. 20. 2º), en los supuestos en los que no se pueda facilitar "información clara y adecuada" en razón de "secretos comerciales e industriales".

5. El uso de la IA por el sector público, transformación y garantías del Derecho Público

5.1. Del incipiente uso público de la IA a un posible deber de uso

La IA sin duda puede usarse en el sector público para innovar y mejorar la eficiencia, la eficacia y los tiempos, así como especialmente la calidad técnica y jurídica de las actuaciones administrativas y la prestación de servicios públicos. Es más, puede anticiparse a las necesidades de cada sujeto para prestarle servicios precisos. Los algoritmos y la IA ya se usan para gestionar movilidad, sostenibilidad, servicios sociales, educativos, de salud, en la contratación de bienes y servicios, o en la selección de personal, en la asignación de subvenciones y beneficios sociales; seleccionan objetivos dónde hacer inspecciones, revisiones o lucha contra fraude y corrupción. En España, con apoyo de conocidos expertos como Falcciani la valenciana Ley 22/2018, de 6

de noviembre, regula exhaustivamente el uso de la tecnología para la prevención y lucha de la corrupción, irregularidades y malas prácticas. También en España, el sistema *VeriPol* estima la probabilidad de que una denuncia sea falsa. O el Sistema *VioGén* señala la peligrosidad de posibles hombres maltratadores y acaba determinando las decisiones de prisión preventiva. El 70% de estados EEUU utilizan desde 2004 algún instrumento mecánico para la decisión de libertad condicional (ej.: *Level of Services Inventory- Revised*, LSI-R). Los sistemas de predicción de riesgo (*risk assessment instruments, Public Safety Assessment*, PSA) después de la condena se emplean en más de veinte jurisdicciones de EEUU en los últimos años. Algunos de estos sistemas se usan por jueces para determinar la condena por la posible reincidencia (sistema COMPAS del caso *Loomis* que luego se comenta). En Argentina, *Prometea*, del Ministerio Público Fiscal de la Ciudad de Buenos Aires, al parecer adopta resoluciones en 15 segundos con un 98% de acierto.

Ahora bien, del lado negativo, ya se han detectado errores masivos en asignación de ayudas públicas (como el sistema *CalWIN* en California o en Texas o el CBMS de Colorado (*Colorado Benefits Management System*). Unos 1.500 pasajeros semanalmente no pueden subir a un avión al ser incluidos erróneamente en la lista "No Fly" o excluidos de contratación pública (Citron, 2009: 1256). En EEUU, la sentencia *K.W. V. Armstrong* 89 F.3D 962, 976 (9TH Cir. 2015) revisó el sistema del *Department of Health and Welfare* que atribuía beneficios de seguridad social por enfermedad en Idaho y algoritmo detectó diferencias desproporcionadas entre distintas zonas y errores estadísticos en la fórmula utilizada. En Francia está cuestionado y pendiente de recursos el sistema de admisión *post-bac* que distribuye estudiantes en establecimientos de educación superior con base en los criterios previstos en el artículo L. 612.3 del Código de educación.

Aún parece relativamente incipiente el uso público de la IA. Es difícil identificar y controlar los usos que están en marcha o van a ser implantadas en un futuro inmediato. Resulta importante la publicidad activa de estas iniciativas y proyectos para su categorización, análisis y evaluación.

Respecto de los usos jurídicos de la IA, cada vez se desarrollan sistemas para la identificación, recuperación y tratamiento de la documentación jurídica de asesoría jurídica, para predecir y apoyar las decisiones jurídicas, para la creación de documentos legislativos, para

la argumentación y negociación jurídicas; se desarrollan sistemas que actúan como consultores o asistentes inteligentes para encontrar solución legal a problemas. La automatización y la IA puede hacer más previsible la aplicación de la ley, garantizar una mayor objetividad y legalidad públicas e incluso reducir la conflictividad y recursos. En esta dirección, cabe mencionar los sistemas de revisión asistida por tecnología o codificación predictiva (*Technology-assisted review*, TAR), son sistemas con aprendizaje automático que permiten buscar y clasificar entre millones de documentos jurídicos los que son relevantes para el supuesto específico. Y ha habido una interesante transición: de la inicial admisión judicial del uso de estos sistemas (en EEUU, Irlanda o Reino Unido), se ha llegado a que se consideren plenamente idóneos, al punto de que los tribunales ya casi imponen su uso. Así, en 2012 un Tribunal en EEUU admitió su uso en el proceso, pues pese a los errores del TAR, la búsqueda humana es más imperfecta (*Da Silva Moore v. Publicis Groupe et al*. S.D.N.Y. (2012). En *Rio Tinto v. Vale*, S.D.N.Y (2015), se afirma ya que el derecho al uso del TAR está totalmente asentado ("*black letter law*"). Y en *Hyles v. New York City* S.D.N.Y. (2016), los tribunales expresan el deseo de que se utilizara el TAR por el poder público. No obstante, se afirma que no se puede –aún- obligar a la ciudad de Nueva York a usa estos sistemas.

Ello nos lleva a pensar que en un futuro nada lejano, el debate no se ceñirá a las garantías y requisitos del uso público de la IA, sino a la obligatoriedad de su incorporación en diversos ámbitos. Y en una línea similar, considero posible que también en algunos ámbitos se deban incorporar garantías, como la necesidad de que se tenga que justificar por qué la decisión pública –humana- no sigue la propuesta elaborada por la IA.

5.2. Cambios conceptuales estructurales en el Derecho Público y la relevancia de los tipos de IA que se empleen

El uso público de sistemas de IA somete al Derecho público a cambios conceptuales estructurales. Es posible que se considere al sistema de IA como nueva fuente del Derecho, al menos para proyectar garantías del procedimiento normativo. En Italia La sentencia de 22 de marzo de 2017 del Tribunal Administrativo Regional de Lazio (Sección III bis), recurso 11419 del 2016 categoriza el algoritmo y software empleado como acto administrativo. Se argumenta que con el software toma

forma la voluntad administrativa y constituye, modifica o extingue las situaciones legales individuales. El software termina identificándose y realizando el mismo procedimiento. Además se afirma que el software utiliza lenguajes generalmente incomprensibles para el funcionario y para el interesado. Además, por lo general lo procesa una empresa privada con el que lo contrata la Administración. De ahí que se considere acto administrativo para atribuir garantías al ciudadano.

También habrá que valorar qué va a constituir el "expediente" o el "procedimiento" y las garantías que corresponden. Hay que examinar la vigencia de las teorías del órgano y el análisis de la real capacidad del titular del órgano y de los humanos de adoptar la decisión, y especialmente, de acceder, comprender, supervisar y controlar el correcto funcionamiento de los sistemas de IA y el cumplimiento de derechos y garantías.

En Australia, el artículo 64 de la *Social Security (Administration) Act 1999* (seguido por otras normas) legitima de modo sencillo, quizá excesivamente, la toma de decisiones automatizadas: "El Secretario puede disponer el uso de programas de ordenador para tomar decisiones [... y en estos casos, la decisión del programa] se considera tomada por el Secretario ".

> *"6A Secretary may arrange for use of computer programs to make decisions*
>
> *(1) The Secretary may arrange for the use, under the Secretary's control, of computer programs for any purposes for which the Secretary may make decisions under the social security law.*
>
> *(2) A decision made by the operation of a computer program under an arrangement made under subsection (1) is taken to be a decision made by the Secretary.*"
>
> Fuente: https://www.legislation.gov.au/Details/C2018C00514
>
> Una lista de otras 28 fuentes son consultables en:
>
> https://airtable.com/embed/shrpkHgfDpvec6BA3/tblHPWVuiNI6v63nn?backgroundColor=blue

La IA ha de afectar a la contratación pública, pues se debe garantizar que los productos o servicios de IA que se demanden por el sector público se sometan a test previos, que las ofertas y pliegos garanticen el cumplimiento legal y la igualdad de las partes que concurran (Joh,

2017). Incluso puede valorarse la obligación de contratar sistemas de código abierto (Citron, 2007: 1308-9). De igual modo, resulta importante el análisis y modelo de configuración de la propiedad y titularidad de los sistemas o servicios de IA que generen y que se contraten por sector público. De ello puede depender el control real del cumplimiento normativo.

Asimismo, va a resultar esencial determinar el régimen jurídico de los macrodatos públicos y privados y las posibilidades de su uso público y libre circulación y el régimen de protección de los mismos y en su caso su consideración como bienes comunes. A este respecto hay que prestar especial atención al Reglamento (UE) 2018/1807 del Parlamento Europeo y del Consejo, de 14 de noviembre de 2018, relativo a un marco para la libre circulación de datos no personales en la Unión Europea.

Las actuaciones automatizadas y el uso de la IA en el sector público puede variar y mucho. En principio, tendrá menos problemas jurídicos el uso de sistemas de IA deterministas según se dijo, en los que la decisión automatizada está determinada a priori por reglas y algoritmos, que por complejos que sean son claros y controlables. Mayores problemas jurídicos se dan respecto de las garantías de motivación, transparencia y otras garantías respecto de los sistemas de aprendizaje automáticos o expertos no deterministas, que van aprendiendo continuamente de los datos, entorno y sus propias actuaciones anteriores.

La introducción de la IA parece más compleja cuando el Derecho confiere a mayor discrecionalidad a la actuación administrativa. De un lado, en principio habrá mayor dificultad técnica de reproducir procesos más complejos de conocimiento y decisión humana. Y a mayor discrecionalidad, jurídicamente debe darse una mayor garantía de que la decisión sea humana y se incrementen las garantías y posibilidades de revisión de la actuación automatizada. Del lado contrario, los actos más reglados, más parametrizables y menos discrecionales son los iniciales candidatos al uso público de la IA, como viene sucediendo en las últimas décadas.

Tiene gran relevancia jurídica el tipo de actuación de la IA, en razón de si el sistema adopta la decisión o simplemente formula propuestas o introduce elementos informativos para que se adopte la decisión por el humano. Es decisivo determinar el grado de automatización de la decisión y en su caso de intervención efectiva del humano en su revisión

y en la decisión adoptada. De ello puede depender que se aplique el derecho a no ser sometido a decisiones individualizadas ya analizado. En sentido paralelo, puede que la actividad automatizada pública ni se consideren como tal por no reunir los requisitos de ser "íntegramente a través de medios electrónicos" y que "no haya intervenido de forma directa un empleado público" (art. 41. 1º Ley 40/2015 española). Si no se dan estos requisitos, no se aplicará el régimen de las actividades automatizadas (por ejemplo, la obligación de determinar el órgano responsable y que éste fije "especificaciones, programación, mantenimiento, supervisión y control de calidad y, en su caso, auditoría del sistema de información y de su código fuente", art. 41. 2º).

Muy buena parte de los usos públicos de IA que se conocen formalmente no adoptan la decisión, que al menos formalmente es humana. En los ámbitos que deben quedar reservados a la discrecionalidad administrativa y a la decisión humana, deben generarse contextos que faciliten que el humano no siga automáticamente la decisión del algoritmo.

5.3. Requisitos y garantías frente al uso público de la IA

Los sistemas inteligentes que se utilicen en el sector público deben contar con garantías en la aprobación y verificación, con especial cuidado de la "traducción" algorítmica que se hace de las normas, conceptos y prácticas jurídicas. Especial importancia adquiere el deber de motivación y el paralelo derecho a una buena administración, el debido proceso y la capacidad de impugnar, alegar y probar frente a las actuaciones a través de algoritmos y aprendizaje automáticos y las garantías y derechos.

En Francia el artículo 4 de la Ley n ° 2016-1321 del 7 de octubre de 2016 para una República digital (*sur la république numérique*) ha regulado los elementos básicos de las decisiones algorítmicas administrativas. Y los mismos han sido desarrollados por el artículo R. 311-3-1-2 del Decreto del 16 de marzo de 2017. El Consejo Constitucional ha revisado dicha legislación en su Decisión N ° 2018-765 DC, 12 de junio de 2018 (nº 70) y somete a tres condiciones el "mero uso de un algoritmo para basar una decisión administrativa individual":

- (1º) la decisión explícitamente ha de señalar que ha sido adoptada sobre la base de un algoritmo. Y a petición del interesado deben

comunicarse las principales características de la implementación del algoritmo. El reglamento precisa que se ha de detallar el grado y modo de contribución del procesamiento algorítmico a la toma de decisiones; los datos procesados y sus fuentes; los parámetros de tratamiento y, en su caso, su ponderación, aplicados a la situación de la persona interesada y las operaciones realizadas por el tratamiento. Asimismo cabe recordar que el artículo 6 de la Ley obliga a publicar en línea las reglas que definen los principales tratamientos algorítmicos utilizados en sus decisiones.

- (2º) La decisión debe ser recurrible administrativamente.

- (3º) El Consejo Constitucional confirma también como requisito la prohibición específica de decisiones exclusivamente automatizadas si incluyen el uso de datos especialmente protegidos. Es más, el nuevo artículo 10.1º Loi n° 78-17 de 6 de enero de 1978 prohíbe las decisiones semiautomatizadas en el ámbito de la justicia.

Especialmente en el continente americano el debido proceso (*due process*) brinda las garantías de la producción de las decisiones públicas judiciales y administrativas. Y para el uso de la IA y el big data en EEUU se ha afirmado el data due process (Crawford & Schultz, 2014: 117 y ss.; Citron, 2009). Consistiría básicamente en un nuevo modelo de auditorías estructuradas de control de calidad, una técnica de gestión sistematizada para descubrir errores, identificar causas, generar incentivos para detectarlos y corregirlos. Se trataría especialmente de un derecho a auditar los datos utilizados para realizar la decisión administrativa o judicial en cuestión. Debe haber una información mínima a los afectados sobre qué predijo el algoritmo, a partir de qué datos y la metodología empleada. En razón del debido proceso se debe poder explicar la fundamentación de la decisión para poder mitigar los posibles sesgos.

Además de garantías de explicabilidad y transparencia que luego se concretan, el debido proceso en principio garantiza la audiencia debida ante un juez imparcial así como garantías para confrontar pruebas y presentar testigos. No obstante, a este respecto se alerta de que respecto las decisiones automatizadas son casi nulas las posibilidades de ser realmente oído en la práctica. Y de que los mismos tribunales suelen ser "deferentes" a las explicaciones de la Administración, especialmente cuando involucran juicios expertos complejos. De-

mostrar que una máquina se equivoca es difícil para el ciudadano y es necesario aumentar sus garantías procesales para obtener, proponer y presentar pruebas.

La proyección de las garantías del debido proceso en cada caso concreto es variable. A más afectación de derechos o intereses individuales, mayor obligación de cumplir con los requisitos del due process. Pues bien, en caso de uso público de algoritmos y big data este balance debe ser "recalibrado" (Citron, 2007: 1286) y no sólo se debe analizar el caso concreto, sino los miles o millones de decisiones que potencialmente seguirán produciéndose erróneamente. Además, ha de tenerse en cuenta que si el error no se revisa, las decisiones pasarán a ser big data que alimentará a los futuros algoritmos haciendo que el sesgo se multiplique.

6. La transparencia explicabilidad, trazabilidad y auditabilidad del uso público y privado de la IA

La transparencia algorítmica genéricamente incluye el acceso a funcionamiento y resultados, explicabilidad, trazabilidad y auditabilidad del uso público y privado de la IA. Como se vio, la transparencia se configura como la clave de bóveda y premisa de los principios éticos y garantías jurídicas. Desde el punto de vista jurídico, las obligaciones de transparencia surgen por diversas otras obligaciones constitucionales y legales.

- La *accountability* y el principio democrático obligan a que se pueda controlar la actuación de programadores, funcionarios y analistas de bajo nivel que "traducen" la ley en un código, muchas veces sin capacidad técnica ni jurídica para hacerlo.
- Respecto de los usos públicos de la IA, se incluyen tanto obligaciones de publicidad activa como el derecho de acceso a la información pública, como los son los programas, algoritmos y sistemas de IA y a los registros que deja la actuación automatizada.
- El régimen de protección de datos incluye fuertes obligaciones de transparencia y de derecho de acceso, las cuales pueden incrementarse cuando se tratan datos especialmente protegidos. Además, el nuevo derecho respecto de las decisiones

automatizadas (art. 22 RGPD-UE) también incluye especiales obligaciones de información que han sido arriba detalladas. Las autoridades de protección de datos tendrán particulares facultades de conocimiento de los tratamientos.

- La transparencia también es esencial para poder controlar errores, discriminación y sesgo algorítmicos públicos y privados. Además, respecto del uso de la IA en el ámbito sancionador, policial y judicial, además, será especialmente importante la transparencia para garantizar la individualización y que se tiene en cuenta la "totalidad" de circunstancias.

- La transparencia en el uso público de la IA tiene una especial conexión con las garantías de la actuación administrativa y el debido proceso. Asimismo, los particulares afectados por las actuaciones tendrán además un reforzado derecho de acceso. Los interesados deben poder conocer la motivación de las decisiones para poder controlar que no son arbitrarias sino ajustadas a Derecho y poder recurrirlas. Hay que partir, en cualquier caso, de que lo importante no es saber lo que pasa por la cabeza de un juez o de funcionario cuando adopta su decisión, sino que la decisión cuente con motivación suficiente y ajustada a Derecho.

Como se ha adelantado, en EEUU en razón del *data due process*, los sistemas automatizados deben generar "audit trails" para cumplir con la obligación de registros para el control judicial que exige el due process Citron (Citron, 2007:1298). Estos audit trails deben seguir las mejores prácticas de la industria y auditores especializados deben poder revisarlos. Los registros deberían incluir el historial de la actividad automatizada (Citron, 2007: 1305); los factores que el algoritmo utiliza y la precisión histórica de los resultados del algoritmo; los porcentajes explícitos de cada factor respecto de cada resultado que arroja o cuanto menos los factores determinantes. En el caso de los sistemas de aprendizaje automático no deterministas, su naturaleza de caja negra hace casi imposible conocer la importancia de cada variable o los efectos finales en las predicciones. No obstante, en algunos casos sí que es posible tener en cuenta los factores tenidos en cuenta en la decisión y se puede controlar si la precisión es suficiente. El problema fundamental puede ser que los datos fundamentales están en manos privadas y lo que se contratan son los resultados, no el algoritmo (Joh, 2017: 18, 19).

El alcance y las facultades de acceso pueden ser muy variable. En general será más respecto de los usos públicos de la IA. Asimismo variará según se trate del conocimiento por la ciudadanía en general, sectores especializados (académicos, industria, asociaciones de consumidores, etc.) o los colectivos o particulares interesados.

Como se afirmó con relación a las autoridades de control, especialmente y cuanto menos, deben reconocerse fuertes facultades de acceso y control a las autoridades judiciales o autoridades independientes correspondientes. Además de los variados sectores en los que se proyecte la IA y las correspondientes autoridades, tendrán especial protagonismo las autoridades de protección de datos y transparencia.

Cabe destacar, por último, algunas resoluciones relevantes de la transparencia de los usos públicos de IA. Así, en Francia, la Comisión francesa *d'accès aux documents administratifs* en su N°1508951/5-2 10 de marzo de 2016 obligó a la *Direction générale des finances publiques* a hacer público el código fuente del programa de ordenador usado para calcular o el impuesto sobre los ingresos de las personas físicas . Y en su *Avis* 20180276 de 19 de abril de 2018 obligó a hacer accesibles los códigos fuente de tres programas que desarrollan modelos con datos económicos usados por el Ministerio de Economía. El órgano de transparencia es favorable a considerar como documentos accesibles el sistema informático y sus algoritmos. Asimismo, pese a que el software esté en constante evolución, ello no impide que "cada versión del código fuente del mismo programa informático tiene el carácter de documento administrativo completado y se puede comunicar" (Resolución de 2016).

En España la Resolución de 21 de septiembre de 2016, de estimación de las Reclamaciones 123/2016 y 124/2016 (acumuladas) de la Gaip, la Comisión de Garantía del derecho de acceso a la información pública de Cataluña. Se afirma que "El código fuente de un programa informático utilizado por la Administración en la designación de los miembros de tribunales evaluadores constituye información pública" y "no parece que la entrega de su código fuente pueda representar un peligro para la seguridad pública". "Existe un interés público y privado [...] en poder comprobar que el programa informático está correctamente diseñado para garantizar la igualdad de todos los participantes y que la designación de los miembros de los tribunales ajusta a los criterios establecidos por la normativa que los regula."

La ya mencionada sentencia de 22 de marzo de 2017 del Tribunal Administrativo Regional de Lazio considera al programa empleado acto administrativo y reconoce un acceso "más profundo" para su completo conocimiento y comprensión por la ciudadanía y por lo interesados.

En EEUU ha habido diversas sentencias relevantes sobre la posibilidad de acceso a las tecnologías utilizadas en el policial y judicial. Es de interés la decisión en el caso *State v. Andrews* 134 A.3d 324 (Md. App. 2016). La policía de Baltimore estaba bajo un acuerdo de confidencialidad impuesto por el FBI y el Fiscal del Estado de Baltimore respecto de la tecnología de *StingRay* de Harris Corporation para vigilancia de teléfonos. El acuerdo de confidencialidad prohibía dar información incluso en los procedimientos judiciales. La decisión judicial considera que el secreto es contrario a los principios constitucionales y el tribunal debe entender cómo se lleva a cabo la vigilancia, la funcionalidad y la gama de información que maneja. Las sentencias han sido contrarias a la transparencia respecto de los casos de análisis ADN probabilístico (como el programa informático *TrueAllele* de Cybergenetics); el sistema facilita datos estadísticos de si el acusado participó en el crimen. Frente a lo que sucede en el ámbito de la IA, en estos casos de ADN probabilístico los tribunales parten de un consenso científico de validez y aceptación generalizada por la comunidad científica. Es por ello que, pese a que el uso de la tecnología afecta a la decisión judicial penal, las sentencias entienden que no es preciso divulgar el programa porque "causaría un daño irreparable a la compañía". Así, *Commonwealth* v. *Foley,* 38, A. 3d 882 (Pa. Super. 2012 (en banc), at 888) y *Commonwealth of Pennsylvania* v. *Michael Robinson,* Memorandum Order, CC201307777, 4 de febrero de 2016.

Sin embargo, para el caso concreto del uso de sistemas predictivos de IA y además en el ámbito penal, resulta sin duda una referencia la sentencia *State v. Loomis*, 881, N.W.2d 749, 7532 (Wis, 2016). En un caso de aplicación de algoritmos predictivos que determinaron el incremento de la pena por la peligrosidad, el tribunal consideró suficiente la información facilitada a la defensa –poco más que un manual de la aplicación- al tiempo que señalaba que los propios jueces garantizaban suficientemente los derechos de la defensa en la aplicación concreta de la herramienta algorítmica.

El proyecto *Muckrock* con el Rutgers Institute for Information Policy and Law hace un importante seguimiento sobre la transparencia en el uso público de algoritmos.

Bibliografía

AEPD - ISMS Forum (eds.); Carlos Alberto Sáiz (coord.). (2017). *Código de buenas prácticas en protección de datos para proyectos de Big Data*, mayo, AEPD e ISMS Forum, Madrid.

AI-HLEG (2018 y 2019). *Draft Ethics Guidelines for Trustworthy AI. Working Document for stakeholders' consultation*, Bruselas, 18 de diciembre 2018 y *Ethics Guidelines for Trustworthy AI. Working Document for stakeholders' consultation*, Bruselas, 8 de abril de 2019.

AI4People FLORIDI, Luciano. *et al.* (2018). "*AI4People* —An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations", *Minds and Machines* 28(4): 689-707. Acceso en https://doi.org/10.1007/s11023-018-9482-5 y https://link.springer.com/article/10.1007/s11023-018-9482-5

CALO, Ryan. (2017). "Artificial Intelligence Policy: A Primer and Roadmap". *University of California, Davis* [Vol. 51:399-435], p. 24 y ss. Citas de la p. 25 (agosto 2017). Acceso en SSRN: https://ssrn.com/abstract=3015350 or http://dx.doi.org/10.2139/ssrn.3015350

CITRON, D. K. (2008). "Technological Due Process", 85 *W*ash. *U. L. R*ev. *1249*: http://openscholarship.wustl.edu/law_lawreview/vol85/iss6/2

COTINO HUESO, Lorenzo (2019 a). "Riesgos e impactos del big data, la inteligencia artificial y la robótica y enfoques, modelos y principios de la respuesta del Derecho", en Boix Palop. A. y Cotino Hueso, L. (coords.), *Monográfico Derecho Público, derechos y transparencia ante el uso de algoritmos, inteligencia artificial y big data RGDA Iustel*, nº 50, febrero 2019. Acceso en https://www.iustel.com/v2/revistas/detalle_revista.asp?id=1

COTINO HUESO, Lorenzo (2019 b). "Ética en el diseño para el desarrollo de una inteligencia artificial, robótica y big data confiables y su utilidad desde el derecho" en *Revista Catalana de Derecho Público* nº 58 (junio 2019) acceso en http://revistes.eapc.gencat.cat/index.php/rcdp/issue/archive

CRAWFORD, K. & SCHULTZ, J. (2014), Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms. *Boston College Law Review, Vol. 55, No. 93, 2014; NYU Law and Economics*

Research Paper No. 13-36. Retrieved from: https://ssrn.com/abstract=2325784.

DILEMATA (2017) *Monográfico Ética de datos, sociedad y ciudadanía*, Núm. 24. Acceso completo en https://www.dilemata.net/revista/index.php/dilemata/issue/view/25

G29-UE. (2018). *Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679*, 3 de octubre de 2017, versión final 6 de febrero de 2018, Doc WP251rev.01

JOH, E. E. (2017). The Undue Influence of Surveillance Technology Companies on Policing (February 27). *N.Y.U. L. Review Online*. https://ssrn.com/abstract=2924620

MAYER-SCHÖNBERGER, Viktor y Cukier, Kenneth (2013 a): *Big Data: A Revolution That Will Transform How We Live, Work, and Think;* ahora en *Big data. La revolución de los datos masivos*, Turner Publicaciones.

O'NEIL, Cathy. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown. Ahora en español (2017). *Armas de Destrucción Matemática. Como el Big Data aumenta la desigualdad y amenaza la democracia*. Capitán. Swing, Madrid, 2017.

PETIT, Nicolas. (2017). *Law and Regulation of Artificial Intelligence and Robots - Conceptual Framework and Normative Implications* (March 9, 2017). Working paper.SSRN: https://ssrn.com/abstract=2931339 or http://dx.doi.org/10.2139/ssrn.2931339

STANFORD UNIVERSITY (2016). *Artificial intelligence and life in 2030. One Hundred Year Study on Artificial Intelligence (AI100)*, Study Panel (Peter Stone, chair), septiembre acceso en https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai100report10032016fnl_singles.pdf

White House. (2016). "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights," *Executive Office of the President*, mayo 2016.