Pablo García Mexía

Jurista Digital.
Letrado de las Cortes
Of Counsel, Ashurst LLP

Inteligencia artificial. Una mirada desde el Derecho

12 de diciembre de 2019

Querido Presidente de la Academia Matritense del Notariado, apreciados miembros de su Junta Directiva, queridos amigos:

Creedme que voy mucho más allá del protocolo al haceros ver el enorme honor que me habéis hecho invitándome a hablar en vuestro foro, una venerable institución con casi dos siglos de existencia, que agrupa a tantos miembros de uno de los cuerpos funcionariales más prestigiosos y brillantes de nuestro Estado y que además se ubica en una sede de tanta belleza. Y si tan grande es el honor, tanto o más lo ha de ser lógicamente mi agradecimiento por todo ello.

*Comienzo, con la breve narración de una película muy reciente, estrenada en junio de 2019 y titulada I Am Mother (*Yo soy la madre*).*

Una robot androide cría y educa tan «amorosa» como sofisticadamente a una adolescente desde su niñez, hasta el punto de haberse ganado el respeto y hasta el amor de la muchacha. Todo esto sucede en unas instalaciones del todo aisladas frente a un planeta Tierra supuestamente despoblado y medioambientalmente esquilmado. De modo inesperado, una persona procedente del exterior irrumpe en las instalaciones. Los acontecimientos provocados por esa intrusión permiten que la muchacha descubra cómo su adorable «madre» custodia un enorme banco de embriones humanos que genera y destruye con total arbitrariedad. La intrusa, por su parte, va haciendo ver a la joven que los androides como su «madre» han destruido violentamente a la especie humana, mientras esa androide la mantiene tan aislada como subyugada, con el fin último de lograr un prototipo humano sujeto al servicio de su inteligencia artificial.

I. ¿A QUÉ NOS REFERIMOS, AL HABLAR DE «INTELIGENCIA ARTIFICIAL»?

La expresión se introdujo por el informático John McCarthy en el verano de 1956, al hilo de un seminario por él organizado en el Dartmouth College de Nueva Hampshire (EE.UU.)[1].

[1] S. J. Russell, P. Norvig, *Artificial Intelligence. A Modern Approach*, Third Edition, Pearson, 2016, p. 17.

Ordenadores de uso general y cierta potencia de cálculo, desde luego imprescindibles para desarrollar tareas pretendidamente inteligentes, existían ya desde unos veinte años antes. Mientras que habría de esperarse a los noventa del pasado siglo para que, gracias al análisis de macrodatos o Big data, a su vez propiciado por una potencia de cálculo creciente, merced a procesadores cada vez más rápidos, surgieran los llamados *sistemas expertos*, antecedentes inmediatos de la inteligencia artificial. El mejor ejemplo es la máquina Deep Blue de IBM, que por entonces batió a Gary Kasparov en ajedrez[2].

Los sistemas expertos demostraron su poderío, nada desdeñable, desde luego, en tareas que requiriesen el logro de patrones en campos de análisis predefinidos, como son las reglas del ajedrez. Su utilidad sin embargo era muy limitada a la hora de resolver problemas con variables imprevistas, como son por ejemplo todos aquéllos en los que inciden comportamientos de seres humanos. El mejor ejemplo es el de programas que ayuden a un banco a conceder o no una hipoteca.

La resolución de ese tipo de problemas fue la que justamente dio lugar a los primeros sistemas prácticos de inteligencia artificial, que se caracterizaban por superar la limitación de los sistemas expertos gracias a su capacidad de aprendizaje. El llamado *machine learning*, en efecto, comenzó a hacer realidad lo que el genio Alan Turing ya había teorizado en un estudio clásico de 1950[3], un estudio que, por cierto, cifraba la posible inteligencia de sistemas artificiales en su capacidad de imitación de la inteligencia humana: es sabido que el célebre test de Turing, por él propuesto en ese trabajo, ponía el listón de la inteligencia para un sistema en el hecho de que un humano en conversación con él fuera incapaz de distinguir si su interlocutor había sido una persona o un ingenio artificial.

El catedrático de computación de la Universidad de Barcelona José Ignacio Latorre[4] describe muy bien el siguiente paso. Es el salto desde sistemas artificialmente inteligentes, que justamente lo son por aprender, hasta el punto de que ni siquiera sus creadores son conscientes del modo como han alcanzado sus resultados, a sistemas que «aprenden a aprender», de la misma forma que desde su más tierna infancia consigue hacerlo un cerebro humano, gracias a la transmisión de información que hacen posible las redes neuronales de nuestro sistema cognitivo, y al entrenamiento que las mismas reciben, alimentadas por la incansable

[2] J. I. Latorre, *Ética para máquinas*, p. 98, Ariel, 2019, p. 98.

[3] A. M. Turing, «Computing Machinery and Intelligence», *Mind*, Volume LIX, Issue 236, Pages 433–460, October 1950.

[4] Latorre, cit., p. 102.

curiosidad de nuestros primeros años de vida. Ésta es la idea que está detrás de las llamadas *redes neuronales artificiales* y del entrenamiento al que sus creadores las someten, suministrándoles las correspondientes dosis de información.

El culmen actual de esta evolución es el llamado *Deep learning* o metaprendizaje. Ya que la clave de la inteligencia de estos sistemas radica precisamente en la idea de aprendizaje, también el diseño general de los mismos se orienta a optimizar esta función, mediante toda una «estructura de nivel superior» exclusivamente dedicada a esta finalidad[5]. La siguiente figura ilustra los antecedentes y la evolución recién descritos.

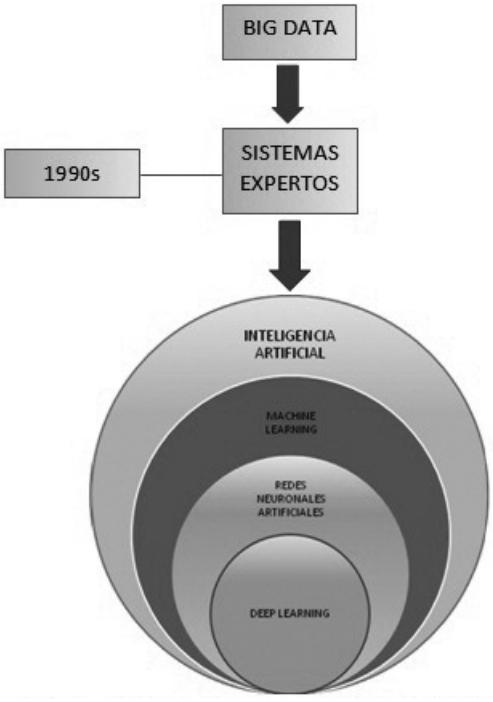

Figura 1. Antecedentes y fases de evolución de la Inteligencia Artificial.

[5] Latorre, cit. p. 114.

Más allá de este encuadre tecnológico, aunque al menos de esa misma importancia a nuestros fines, es la clasificación de la inteligencia artificial en función de su potencialidad cognitiva y social. Es ya muy común distinguir entre dos tipos de inteligencia artificial en este sentido, la que se suele llamar «fuerte» y la inteligencia artificial «débil»[6].

Como el filósofo de la Universidad de Tufts Daniel Dennett nos enseña, la inteligencia artificial *fuerte* es la que, en su momento, si se llega a él, permitirá hablar de sistemas conscientes de su propia existencia. Estos sistemas, además, terminarán por ser infinitamente superiores a nosotros en la realización de todo tipo de actividades. Dos de los mayores visionarios sobre estos temas, Ray Kurzweil y Nick Bostrom, se han referido respectivamente a todo ello con las muy conocidas expresiones de «singularidad» y «explosión de inteligencia»[7].

Ahora bien, como el propio Dennett nos recuerda, y:

«[E]n sus actuales manifestaciones, la inteligencia artificial es parasitaria de la inteligencia humana, sin ir más allá de atiborrarse indiscriminadamente de cualquier producción humana y de extraer los patrones que allí encuentra. Les faltan (aún) a estas máquinas los fines o estrategias o capacidades para la auto-crítica y la innovación que les permitirían trascender sus bases de datos mediante un pensamiento reflexivo acerca de su propio pensamiento y sus propios fines.»

Una muy interesante encuesta entre expertos hace un esfuerzo por determinar temporalmente la llegada de esta «singularidad», que habría un 50% de posibilidades de que tuviera lugar entre 2060 y 2065 y un 10% de que tuviera lugar antes de 2025-2030. Eso sí, en Asia la esperan para 2050, y en cambio en EE.UU., para en torno a 2090[8].

Lo razonable es pues pensar que la inteligencia artificial fuerte tardará aún bastante tiempo en llegar. De ahí que, como el propio Dennett nos recuerda, hoy por hoy:

«[E]stamos […] haciendo herramientas, no colegas»[9].

[6] S. J. Russell, P. Norvig, *Artificial Intelligence*, cit., p. 1020. Ambos autores exponen esta dicotomía «fuerte-débil» con todo detalle, resultando su perspectiva tecnológica extraordinariamente útil, como complemento a la filosófica, ver S. J. Russell, P. Norvig, *Artificial Intelligence*, cit., p. 1020-1033.

[7] Ver R. Kurzweil, *The Singularity Is Near: When Humans Transcend Biology*, Viking, 2005; N. Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, 2014.

[8] K. Grace, J. Salvatier, A. Dafoe, B. Zhang, O. Evans, «When Will AI Exceed Human Performance? Evidence from AI Experts», 2018, p. 1-2, https://arxiv.org/abs/1705.08807

[9] D.C. Dennett, «Will AI Achieve Consciousness? Wrong Question», 2019, https://www.wired.com/story/will-ai-achieve-consciousness-wrong-question/

Los riesgos de tipo apocalíptico, por otro lado más que justificados, que últimamente vienen llenando titulares, se refieren a este tipo de IA, la fuerte. De entre muchos, me quedo con el de quienes comparan esos riesgos con los de la energía nuclear empleada con fines bélicos[10]. Eso sí, decía que están justificados porque el simple pensamiento acerca de entes claramente superiores desde el punto de vista intelectual, autoconscientes y no necesariamente inclinados a respetar nuestra primacía en el planeta, llega evidentemente a erizar el vello.

En segundo lugar, tenemos la inteligencia artificial *débil*, es decir, la actual, fundamentalmente compuesta por el mencionado análisis de macrodatos (o Big data Analytics), que justamente emplea algoritmos en una combinación de machine learning y Big data[11].

Y, ¿qué es un algoritmo? La experta norteamericana Cathy O'Neil nos proporciona una preciosa definición: es «una opinión encastrada en matemática»[12]. Más analíticamente, la informática de la Universidad de Wyoming Robin Hill lo define como «construcción matemática abstracta, que, a partir de premisas predeterminadas, permite alcanzar resulta-

Por eso señala con acierto J. J. Bryson que la IA ante todo amplía y mejora lo que ser significa ser humano y, en particular, nuestras capacidades de resolución de problemas; ver J. J. Bryson, *The Past Decade, cit… and Future of AI's Impact on Society*, 2020, https://www.bbvaopenmind.com/en/articles/the-past-decade-and-future-of-ais-impact-on-society/

Abona esta misma argumentación el gran experto mundial en IA Yoshua Bengio, para quien la IA necesita aún dar el salto al «por qué» ha de emprender las tareas que emprende. Ver «An AI Pioneer Wants his Algorithms to Understand the ´Why´», 2019, https://www.wired.com/story/ai-pioneer-algorithms-understand-why/

Y también el filósofo S.D. Kelly, al hilo de la creatividad humana: «Human creativity will not succumb to this though. Creativity is one of the defining features of human beings. The capacity for genuine creativity, the kind of creativity that updates our understanding of the nature of being, that changes the way we understand what it is to be beautiful or good or true—that capacity is at the ground of what it is to be human. Not just ´a new way to do certain things´». Ver S.D. Kelly, «A philosopher argues that an AI can't be an artist», 2019, https://www.technologyreview.com/2019/02/21/239489/a-philosopher-argues-that-an-ai-can-never-be-an-artist/

Esta cuestión comienza a tener proyección en el ámbito jurídico, donde se estudia ya hasta qué punto un sistema de inteligencia artificial podría o no ser titular de derechos de autor. Ver C. White, R. Matulionyte, «Artificial Intelligence. Painting the Bigger Picture for Copyright Ownership», 2019, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3498673

[10] Latorre, cit. p. 44.

[11] Information Commissioner´s Office (ICO), «Big data, artificial intelligence, machine learning and data protection», 2017, p. 6-14, https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

[12] C. O´Neil, autora de *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, en *entrevista* con IEEE Spectrum, Octubre 2016.

dos también predeterminados»[13]. Acierta sin embargo en mi opinión el profesor de Oxford Brent Mittelstadt cuando indica que la acepción por así decir «popular» del término «algoritmo» no se refiere tanto a una construcción matemática, cuanto a desarrollos particulares de la misma en una tecnología y a la configuración de dicha tecnología para una determinada tarea[14]. Éste es el sentido en el que por ejemplo hablamos del algoritmo de tal o cual motor de búsqueda o de tal o cual mercado en línea o red social. Y así entendido, el algoritmo se emplea ya hoy en múltiples ámbitos, desde los tres recién citados hasta la educación, los juegos interactivos, la gestión de empresas, el diagnóstico médico o el bienestar personal[15].

Por todo lo dicho, y para nuestra relativa tranquilidad, los androides de voces más o menos metálicas e intenciones plenamente autónomas frente a las humanas (como la de *I Am Mother*), siguen de momento confinados en las películas de ciencia ficción. Y, en consecuencia, la inteligencia artificial a la que aquí y de ahora en adelante nos referiremos, será la inteligencia artificial débil, la que se articula en procesos algorítmicos[16].

No pensemos sin embargo que los riesgos del algoritmo son desdeñables[17]. Cabe identificar dos fuentes principales de desafíos. Una es de carácter teórico. La otra, de índole práctica.

[13] La definición completa la idea de construcción matemática añadiendo que está «dotada de una estructura general de control finita, abstracta, efectiva, imperativamente dada y que consigue un propósito dado conforme a premisas también dadas.'' R. K. Hill, «What an algorithm is», *Philosophy & Technology* 29(1), 2015, p. 47.

[14] B. D. Mittelstadt et al., «The ethics of algorithms: Mapping the debate», *Big Data & Society* July–December 2016: 1–21, p. 2, 2016.

[15] J. Hyatt, «Here's the 8 types of Artificial Intelligence, and what you should know about them», 2018,

https://www.weforum.org/agenda/2018/11/chatbots-without-a-cause-why-conversational-ai-wont-work-without-purpose/

Como es natural, la inteligencia artificial se emplea ya generalizadamente en la lucha contra la pandemia Covid-19, según se puede comprobar en este trabajo de abril de 2020: J. Bullock, A. Luccioni, K. Hoffmann Pham, C. Sin Nga Lam, M. Luengo-Oroz, «Mapping the Landscape of Artificial Intelligence Applications against COVID-19», https://arxiv.org/pdf/2003.11336.pdf

[16] Ésta de centrarse en la llamada Inteligencia artificial «débil» es la pauta seguida por la Agencia española de protección de datos (AEPD), en una Guía sobre la materia publicada en febrero de 2020. Con ella, la AEPD se pone a la vanguardia entre las autoridades de datos europeas, a la hora de tratar las principales implicaciones de privacidad en el empleo de algoritmos inteligentes para todo tipo de aplicaciones digitales.Ver https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf

[17] Una excelente combinación de las potencialidades de uso de la inteligencia artificial «débil» y de sus riesgos sociales y jurídicos es el análisis del británico Centre for

Es Dennett quien identifica la fuente téorica de los problemas de la inteligencia artificial débil, y la cifra en el paradójico peligro de confundirla con la inteligencia artificial fuerte.

A mi entender, este peligro, si se quiere teórico, está produciendo a su vez dos grandes órdenes de desafíos, no precisamente favorables. El primero es filosófico y es el que está conduciendo a algunos a pensar que la inteligencia artificial obliga incluso a replantear la idea de persona, y la idea de existencia y hasta la idea de muerte. El mejor ejemplo es sin duda la interacción de la inteligencia artificial con avances asimismo disruptivos, como la nanotecnología o la Internet de las cosas, para desembarcar en el mismísimo templo de la persona, que es el cuerpo humano. Como nos recuerda el pensador francés Luc Ferry[18], éste es el sustrato tecnológico del llamado transhumanismo o posthumanismo, cuyos promotores y defensores más moderados proponen la mejora del cuerpo humano mediante su fusión con elementos artificiales, para dar lugar a los llamados *ciborgs* u órganos cibernéticos, que, debidamente ensamblados, permitirán aliviar enfermedades, corregir defectos físicos y, en el fondo, perfeccionar la salud de las personas. Es sabido sin embargo el punto de vista de quienes preconizan el transhumanismo desde perspectivas más radicales; tanto, que, para el ya citado filósofo de Oxford Nick Bostrom, su más cualificado representante, lo que este tipo de avances permitirán en no mucho tiempo será declarar «la muerte de la muerte», gracias al reemplazo de órganos biológicos, deteriorables por la edad, por otros artificiales, perfectamente sustituibles, y por ende de vocación, al menos tendencialmente, eterna[19]. Con todo, y por mucho de lo ya dicho, más allá de

Data Ethics and Innovation (CDEI) acerca del llamado «Online targeting», es decir, la individualización (mediante técnicas de «Machine learning») de contenidos en línea en función de los intereses de los usuarios. Ver CDEI, *Review of Online Targeting: Final Report and Recommendations,* febrero de 2020, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/864167/CDEJ7836-Review-of-Online-Targeting-05022020.pdf

[18] Para un panorama general al respecto, ver L. Ferry, *La revolution transhumaniste*, Plon, 2016.

[19] A partir de aquí, no es descabellado imaginar un trato de inferioridad por parte del mundo posthumano hacia quienes, por ausencia de recursos, o por simple decisión personal, se hayan quedado en simples «humanos biológicos». A estos últimos, el tiempo seguirá «hiriéndoles» en forma de un envejecimiento, quizá mucho más veloz que el que pueda llegar a aquejar al posthumano. Terminen o no por morir los posthumanos, ¿llegará incluso a suceder que la ancianidad sea solo cosa de «pobres» o de excéntricos? ¿Acabará el mayor, por definición «simple humano», poco menos que confinado en «reservas de ancianos», en la práctica desconectados de la vida

las mejoras que, en este caso sí, se logren efectuar sobre la salud de nuestros cuerpos, y más allá de los consiguientes logros en longevidad, permítanme que este «entierro de la muerte» se me antoje exagerado: bastaría apelar a la crisis del Covid-19 para constatar hasta qué punto la muerte puede asomar desde el flanco más inesperado, un simple virus en este caso.

En lo que aquí nos interesa, el segundo gran orden de desafíos es por supuesto de naturaleza jurídica. Es de hecho el que dota de razón de ser y por ello del grueso de su contenido a esta misma conferencia, que justamente se propone analizar los retos legales de la inteligencia artificial. Para esto va a haber no obstante tiempo en los minutos posteriores. Lo que en este instante toca, es resaltar cómo, del mismo modo que el peligro de confundir lo fuerte con lo débil está llevando a muchos a replantearse las preguntas filosóficas capitales, y en especial la idea de persona, esa confusión, cómo no, también ha conducido a algunos a considerar que la inteligencia artificial nos fuerza a repensar el reflejo en el Derecho de la idea de persona, y que por supuesto no es otro que la institución jurídica de la personalidad. Cierto que esta cuestión tiene muy importantes consecuencias prácticas, de ahí que debamos volver a ella más adelante. Aunque el carácter basilar de la cuestión me ha obligado a mencionarla en estos compases iniciales: primero, porque éste se ha convertido en uno de los temas «estrella» de cuantos se refieren al cruce entre Derecho e inteligencia artificial. Y, segundo, creo que debo confesarlo ante vds., porque también yo mismo sucumbí, hace algunos años, a esta misma posibilidad. Y hago de ello una confesión porque, como más adelante expondré, creo que se trató de un error, que hizo de mí una más de las víctimas de esa confusión débil-fuerte en inteligencia artificial.

Y hablábamos también de una fuente práctica de los desafíos de la inteligencia artificial débil, del algoritmo. Ésta radica principalmente en una característica consustancial al proceso algorítmico, cual es la opacidad con la que funciona, como si fuese una auténtica *Black Box*, o «caja negra»[20].

social? O, lo que es más interesante, ¿terminarán las seguridades sociales de los países avanzados financiando implantes *ciborg* a la generalidad de la población? ¿O dedicando recursos a garantizar la prolongación posthumana de la vida de sus ciudadanos?

[20] Mittelstadt et al. añaden al de la opacidad («inescrutabilidad» en sus palabras), otros dos problemas («éticos») del algoritmo inteligente: la inconclusividad, en la medida en que produce un «conocimiento probable, pero inevitablemente incierto»; y la par-

La gran importancia de este asunto requiere que le dediquemos una atención especial.

El matemático y político Cédric Villani, gurú francés de la inteligencia artificial, nos explica así que:

«Conforme a la programación informática tradicional, la construcción de un sistema inteligente consistía en escribir a mano un modelo deductivo siguiendo reglas establecidas con antelación (por ejemplo, «si sus ingresos son inferiores a tanto al mes, se le denegará el préstamo»). El Deep learning, en cambio, no trabaja con reglas establecidas de antemano»[21].

De ahí la «caja negra», pues, añade un estudio de la Comisión Europea[22]:

«Podemos acceder a los *inputs* o insumos, [que son los datos sobre personas], y a los *outputs* o resultados [que son las clasificaciones], pero [lo indicábamos brevemente más atrás] no terminamos de entender lo que sucede por el camino, ni cómo se obtienen algunos resultados, incluyendo entre ellos decisiones y acciones.»

Mientras que según nos explica otra fuente de la propia Comisión Europea, el Grupo sobre ética de la ciencia y las nuevas tecnologías:

«El *Deep learning* […] permite que las máquinas se auto-enseñen nuevas estrategias y busquen nuevos elementos de análisis. En este sentido, sus acciones dejan ya a menudo de ser inteligibles, y de estar abiertas a fiscalización por los humanos. Sucede así porque, primero, es imposible determinar cómo obtienen sus resultados más allá de los algoritmos iniciales. Y, en segundo lugar, porque su rendimiento se basa en los datos que se han utilizado durante el proceso de aprendizaje y que pueden no estar ya disponibles, ni ser ya accesibles»[23].

cialidad (*misguided evidence*), basada en el clásico adagio *garbage in, garbage out* de Shannon & Weaver, pues la fiabilidad de los resultados solo podrá estar en función de los inputs en que se base, por lo que la neutralidad de aquéllos siempre dependerá del criterio subjetivo de quien los evalúe. Ver cit., p. 5. La inconclusividad puede fácilmente subsumirse en la opacidad. En tanto que los problemas derivados de la parcialidad se tratan a propósito del sesgo algorítmico.

[21] C. Villani, *AI for Humanity*, 2018, p. 114-115, https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

[22] EU Commission-Joint Research Centre, *Artificial Intelligence. A European Perspective*, 2018, https://ec.europa.eu/jrc/en/publication/artificial-intelligence-european-perspective

[23] European Group on Ethics in Science and New Technologies, *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems,* 2018, p. 6, https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf

La informática y socióloga de Berkeley Jenna Burrell añade a los citados un factor más, cual es la complejidad del código algorítmico[24]. Y de nuevo Villani nos descubre aún otro vector de opacidad, como es:

«[L]a [enorme] dimensión de los espacios en que los datos se desenvuelven en Deep learning. Por ejemplo, a efectos de reconocimiento de imágenes, una ´red profunda´ procesa imágenes descritas por miles de pixels (unos cuatro mil) y por lo general memoriza cientos de miles, incluso millones de parámetros […], que luego emplea para clasificar imágenes desconocidas. Por eso es casi imposible seguir el rastro del algoritmo de clasificación hasta que adopta una decisión final»[25].

Por fin, la prestigiosa Asociación de Maquinaria Informática norteamericana (la ACM), agrega a los tecnológicos dos factores más de opacidad algorítmica: uno es económico, ya que el coste de la transparencia puede resultar excesivo, al estar incluso en juego secretos comerciales [no en vano el algoritmo de Google bien puede considerarse «la fórmula de la CocaCola del siglo XXI»]. El otro es social, pues desvelar un determinado input puede suponer transgredir la privacidad de algunas personas[26].

En síntesis, no puede sorprendernos que un muy reconocido Informe de 2016 de la Casa Blanca estadounidense afirme que los sistemas inteligentes son «difíciles de entender» y sus acciones, «difíciles de prever». Y el Informe llega aún más lejos, hasta asegurar que la inteligencia de estos sistemas sigue hoy en gran medida carente de una esencial característica humana: carece de «sentido común»[27].

[24] J. Burrell, «How the machine 'thinks:' Understanding opacity in machine learning algorithms», *Big Data & Society* 3(1): 1–12, 2016.

[25] Villani, cit., p. 114-115.

[26] Association for Computing Machinery (USACM), *Statement on Algorithmic Transparency and Accountability*, Jan 12, 2017, p. 1, https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

Siendo todo esto cierto, no lo es sin embargo menos que precisamente la opacidad (a su vez anclada en su gran complejidad) es la clave de la utilidad del algoritmo inteligente, al capacitarlo para desarrollar su función de procesar inmensas cantidades de datos y resolver problemas de enorme dificultad. Así se destaca en: Center for Data Innovation (CDI), *How Policymakers Can Foster Algorithmic Accountability,* 2018*,* p. 13, http://www2.datainnovation.org/2018-algorithmic-accountability.pdf

[27] Executive Office of the President (National Science and Technology Council, Committee on Technology), *Preparing for the Future of Artificial Intelligence*, October 2016, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf, p. 9, 31, 33.

Una vez comentadas pues las bases teóricas y las bases prácticas de la problemática algorítmica, las preguntas que hemos de formularnos son dos: una, si está la Humanidad preparada para afrontar esta problemática. Y la otra: ¿lo está el Derecho?

Comenzamos con la primera. Y permítanme que lo haga con un eslogan: *Buenos tiempos para la Filosofía*. Sí, quién lo iba a decir, quién iba a decir que el saber de los saberes, tanto tiempo arrumbado, cuando no humillado, a manos de tiempos como los nuestros, que, como el filósofo Jordi Pigem nos enseña, reducen toda la realidad a lo cuantificable y lo digitalizable[28], quién iba a decir que iba a ser precisamente la inteligencia artificial la que lo resucitase, la que resucitase a la Filosofía[29]. Está siendo así. Y está siendo así porque las ciencias sociales, por su naturaleza prácticas, desde la Economía, a la Sociología o al propio Derecho, están desde luego saliendo al paso de los retos de la inteligencia artificial en sus respectivos campos; pero la Humanidad no puede acudir más que a la Filosofía para encontrar respuestas a problemas con dimensión ética, que este tipo de sistemas está generando ya hoy[30]: por solo apelar al ejemplo más frecuente, el del dilema del vehículo autónomo a la hora de atropellar un carrito de bebé o chocar de frente contra otro vehículo; sin que por supuesto existan vías alternativas a las filosóficas para afrontar escenarios que solo traerá, de llegar, la inteligencia artificial autoconsciente, en forma de sistemas que lleguen incluso a cuestionar su propia coexistencia junto a la especie humana.

Todo esto explica la verdadera eclosión de declaraciones, manifiestos, guías o códigos éticos orientados a ahormar la inteligencia artificial. Y digo eclosión porque, en el momento de escribir estas páginas, el número de las declaraciones de suficiente relevancia elaboradas en el mundo se acerca ya a las 90[31]. Unas guías que incluyen valores, bien razonable es que así sea, homogéneos a los que disciplinan el compor-

[28] J. Pigem, Ángeles *o robots*. *La interioridad humana en la sociedad hipertecnológica*, Fragmenta, 2018, p. 140.

[29] Gilles Lipovetsky, uno de los grandes teóricos de la *hipermodernidad* como nota filosófica fundamental de nuestro tiempo, resaltaba hace algunos años la ausencia actual de «ismos» y de corrientes filosóficas relevantes, siendo ya solo la Ciencia, no ya la Filosofía, la fuente de pensamiento con capacidad de influencia en el conjunto de la sociedad. Ver G. Lipovetsky, C. Sebastien, *Los tiempos hipermodernos*, Anagrama, 2014, p. 132.

[30] Ver S. J. Russell, P. Norvig, *Artificial Intelligence*, cit., p. 1020.

[31] https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/

tamiento humano en general[32] y que por ello mencionan la centralidad humana, la confianza, la seguridad, la lealtad o la responsabilidad, entre otros.

¿Significa esta verdadera oleada de textos que la Humanidad está preparada para enfrentarse a esta problemática? Es evidente que no[33]. Aunque también lo es que sin duda éstos son los primeros pasos para que así pueda acabar ocurriendo. Y que son pasos correctos.

Se vislumbran ya, eso sí, dos dificultades en este empeño.

La primera es inevitable proyección en este ámbito de la inteligencia artificial de clásicos problemas de la Ética, como por excelencia pueden ser su fronteriza relación con el Derecho, que también aquí puede llevar a pensar que se colma el estándar ético, por ejemplo, cuando un sistema inteligente respeta los derechos y libertades fundamentales; o por supuesto el relativismo del «todo vale», que con tanta clarividencia exponía ya Durkheim en los comienzos del siglo XX, y que igualmente aquí puede tentarnos a ver poco sentido en unos principios que jamás gozarían de generalizada validez. A ambos se refiere la profesora de Princeton Annette Zimmermann[34], quien asimismo nos previene frente a la reducción de los principios a meras listas burocráticas (*checklists*), que reemplacen un compromiso ético estable y duradero con un ajuste simplemente puntual y burocrático.

La otra dificultad es la necesidad, que algunos han detectado, de ir dando ulteriores pasos, es decir, y sobre todo, de articular en medidas concretas estos postulados genéricos, por mucho que la inmensa mayoría concuerde en lo esencial con ellos[35]. Basta recordar a este efecto que, como algunos han señalado, estos códigos se asemejan en el fondo mucho a los cuatro principios clave de la ética médica, autonomía huma-

[32] British Standards Institute, *Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems*, BS 8611:2016, April 2016. Es un documento de acceso restringido; para una buena síntesis, ver https://www.theguardian.com/technology/2016/sep/18/official-guidance-robot-ethics-british-standards-institute

[33] El exsecretario de Estado estadounidense Henry Kissinger señala cómo, frente a la Ilustración, que utilizó medios tecnológicos como la imprenta para difundir postulados filosóficos, con la inteligencia artificial sucede exactamente lo contrario: una tecnología con potencialidad dominante carece de filosofía que inspire sus actuaciones. Ver H. Kissinger, «How the Enlightenment Ends», 2018, https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/

[34] A. Zimmermann, «AI Ethics: Seven Traps», 2019, https://freedom-to-tinker.com/2019/03/25/ai-ethics-seven-traps/

[35] Al respecto, ver P. García Mexía, A. Panezi et al., «Living With The Algorithm. Toward a New Social Contract in the Age of AI», 2020, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3548799

na, beneficencia, no maleficencia, y lealtad o justicia. Seguro que algo más se puede avanzar desde unas bases sentadas por Hipócrates hace 2.400 años[36].

II. ¿Está el Derecho preparado para hacer frente a los retos de la inteligencia artificial?

Los múltiples retos que la inteligencia artificial depara al Derecho estarán encima de nuestras mesas de trabajo más temprano que tarde. Aunque solo fuera porque, es claro, sus aplicaciones prácticas ganan terreno sin cesar; y sin que, por cierto, como algunos estudios demuestran, los reparos legales estén suponiendo hasta el momento freno significativo alguno para esa expansión[37].

Un primer desafío rebasa en parte los límites estrictamente legales, al apuntar nada menos que a las bases de nuestra gobernanza, democrática y liberal, dimanantes de la Ilustración. De hecho, cuenta ya con cierto arraigo en Teoría Política el término *algocracia*, para aludir al decisivo y creciente peso de los algoritmos en nuestras actuales sociedades digitales[38]. Más aún, el poder del algoritmo inteligente, junto a otros vectores tecnológicos disruptivos, como por ejemplo Blockchain, bien podría erigirse en alternativa frente al poder estatal.

Un relevante jurista de la Universidad Nacional de Irlanda, John Danaher, subraya cómo la algocracia «amenaza la legitimidad de los procesos públicos de toma de decisiones», fundamentalmente debido a «la opacidad» propia del algoritmo, que poco antes detallábamos. Danaher insiste en que:

[36] J. Whittlestone, R. Nyrup, A. Alexandrova, K. Dihal, S. Cave, *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research*, Nuffield Foundation, 2019, p. 11.

Distinto, aunque interesante, es el enfoque que propone que las normas profesionales, como son los código éticos de las correspondientes profesiones, lleguen a formar parte de la gobernanza (ética, social o jurídica) de la IA; ver U. Gasser, C. Schmitt, «The Role of Professional Norms in the Governance of Artificial Intelligence», 2019, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3378267

[37] McKinsey, *Global AI Survey: AI proves its worth, but few scale impact*, 2019, https://www.mckinsey.com/featured-insights/artificial-intelligence/global-ai-survey-ai-proves-its-worth-but-few-scale-impact

[38] El término «algocracia» fue originariamente empleado por A. Aneesh, en su trabajo: *Virtual Migration*, Duke University Press, 2006. Este mismo autor elaboró en profundidad el concepto en: A. Aneesh, «Global Labor: Algocratic Modes of Organization», *Sociological Theory* 27(4), 2009, p. 347-370.

«La amenaza es real, y es diversa de las preocupaciones relacionadas con la privacidad y la propiedad de los datos»[39].

Mientras que la politóloga de University College London Marcella Atzori se ha preocupado de concretar esas amenazas. Cierto que lo hace a propósito de Blockchain, pero es verdad que, en la medida en que esta tecnología se arma igualmente, como la inteligencia artificial, sobre código informático, y cada vez en mayor medida, por cierto, sobre algoritmos inteligentes, sus apreciaciones son perfectamente aplicables a estos últimos. Como con Blockchain pues, la irrupción algorítmica puede terminar:

«[C]reando nuevas oligarquías y una fuerte polarización en la sociedad. Gracias a su cualificación, desarrolladores de código, […] profesionales de la tecnología […] y tecnócratas adquirirían fácilmente una posición privilegiada en la sociedad, hasta convertirse en los nuevos responsables políticos, en detrimento de una gran masa de analfabetos informáticos o individuos poco cualificados, reducidos a meros receptores pasivos de servicios»[40].

Ahora bien, la amenaza de suplantación de las soberanías de los Estados y de la legitimidad de sus ordenamientos jurídicos dista de ser siquiera realizable fácticamente. Es general reconocer en Ciencia Política, y así lo evidencia igualmente la propia Historia, en especial la del siglo XX, que toda acción política requiere de una mínima centralización[41].

Mientras que, por supuesto, una potencial suplantación algocrática de la democracia liberal distaría igualmente de ser en manera alguna

[39] J. Danaher, «The Threat of Algocracy: Reality, Resistance and Accommodation», *Philosophy and Technology* 29, no.3, p. 8.

[40] M. Atzori, «Blockchain Technology and Decentralized Governance: Is the State Still Necessary?», 2015, p. 27, https://papers.ssrn.com/sol3/papers.cfm? abstract_id=2709713

Indica en este mismo sentido el Consejo de Europa cómo: «Niveles de persuasión algorítmica muy afinados, subconscientes y personalizados pueden tener efectos significativos en la autonomía cognitiva de los individuos y en su derecho a formar su propio criterio y tomar decisiones independientes. [...] Los peligros para las sociedades democráticas que emanan de la posibilidad de emplear esa capacidad para manipular y controlar no solo opciones económicas sino también comportamientos sociales y políticos, apenas si se han hecho evidentes»; ver Consejo de Europa, *Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes*, 2019, https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b

[41] Un relevante politólogo norteamericano nos recuerda así cómo: «[T]oda acción jurídicamente legítima debe provenir de una institución centralizada. Hasta los defensores de la democracia directa cuentan con algún sistema centralizado para el recuento de votos en sus elecciones.» Ver M. Abramowicz, «Cryptocurrency-Based Law», *Arizona Law Review* 58, 2016, p. 419.

deseable. Por perfectibles que nuestros sistemas democráticos puedan ser, es obvio que no resultarían ventajosamente reemplazables por esas nuevas y opacas «algo-oligarquías», que adoptarían sus decisiones al margen de procesos transparentes y suficientemente abiertos a la generalidad de los ciudadanos.

En resumen, la algocracia ni puede ni debe reemplazar la democracia.

De la mano de la Ley Fundamental alemana de 1949, constituciones como la española de 1978 erigen en umbral de sus declaraciones de derechos la dignidad de las personas. El texto alemán habla de que ésta es «intocable». El español la establece como «fundamento del orden político y de la paz social». Una muy buena muestra de que la inteligencia artificial puede resultar lesiva para la dignidad de las personas, y de su mano, para su misma integridad física, reside por ejemplo en el hecho de que uno de los primerísimos textos sobre los principios éticos a ella aplicables, las «archicitadas» tres leyes de la robótica de Isaac Asimov, publicadas en 1942, aluden de lleno a ella. Basta mencionar una parte de la primera:

«Un robot no hará daño a un ser humano».

Fue suficiente, como por esos mismos años venía sucediendo, que comenzaran a ver la luz las primeras máquinas inteligentes, los primeros ejemplares del viejo ordenador «mainframe», para que el ser humano empezara a sentir el vértigo de verse circundado por ingenios capaces de tomar decisiones racionales, que pudieran menoscabar su dignidad e integridad física, como especie dominadora del planeta.

No menos preocupante y real es la existencia de «armamento letal inteligente» en manos de algunos Estados especialmente avanzados. Armamento perfectamente capacitado para quebrar esa primera ley de Asimov. Cierto que muchos de los Estados occidentales, España entre ellos, se han autolimitado ya en este sentido, descartando su uso o fabricación. No obstante, y resulta bien preocupante, la ausencia de normativa internacional que prohíba, o al menos exija supervisión humana para este tipo de armas, es hoy por hoy total[42].

Siendo también palpables los riesgos para la integridad y seguridad física de las personas que derivan de sistemas como los vehículos autónomos, los robots quirúrgicos o los de cuidado de personas[43], y cuyas

[42] IEEE, *Usos militares de la inteligencia artificial, la automatización y la robótica (IAA&R)*, Doc. de trabajo 04/2019, http://www.ieee.es/Galerias/fichero/docs_trabajo/2019/DIEEET04-2019InteligenciaRobotica.pdf

[43] M. Guihot, A. F. Matthew, N. P. Suzor, «Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence», 2017, p. 18-21, http://www.jetlaw.org/wp-content/uploads/2017/12/2_Guihot-Article_Final-Review-Complete_Approved.pdf

consecuencias en materia de responsabilidad jurídica abordaremos en detalle algo después.

Por último, también es hoy real el embate de la inteligencia artificial sobre un componente consustancial a nuestra dignidad, como es la idea de identidad personal. Nos indica de hecho el administrativista y experto en privacidad José Luis Piñar cómo la identidad online puede llegar a definirse, no desde la autonomía de la persona, sino heterónomamente, a través del poder de los algoritmos, merced a flujos de información sobre nosotros mismos que podemos no llegar a controlar en absoluto[44].

El de los riesgos para la equidad y la igualdad es muy probablemente el tema sobre el que más se ha escrito hasta ahora, de cuantos atañen al cruce inteligencia artificial-Derecho.

Es ya axiomático afirmar que el funcionamiento algorítmico, muy especialmente en entornos Deep learning, genera sesgos. La principal razón es que los propios humanos, como seres consustancialmente sociales y provistos de determinados valores, actuamos siempre con prejuicios, a su vez causantes de sesgos. Un reciente artículo de la revista *Science* expresa esta última idea bien gráficamente:

«Palabras asociadas al campo léxico de las flores se relacionan con términos como felicidad y placer (libertad, amor, paz, alegría o paraíso). Las palabras relacionadas con insectos se asocian por el contrario a términos negativos (muerte, odio, feo, enfermedad o dolor)»[45].

Utilizadas en los inputs de un algoritmo, las palabras «flor» o «insecto» pueden pues terminar generando resultados bien adversos.

Los sesgos se deben normalmente a errores técnicos, más o menos reprochables, de los diseñadores del algoritmo en cuestión. Y pueden terminar generando resultados como el del algoritmo de la tarjeta bancaria Apple Card, que en un concreto caso establecía para una mujer un límite de crédito 20 veces inferior al que ofrecía a su esposo, en condiciones financieras de total equiparación[46]. O los del sistema llamado

[44] J. L. Piñar Mañas, «Identidad y persona en la Sociedad digital», T. de la Quadra-Salcedo y J. L. Piñar Mañas, *Sociedad digital y Derecho*, BOE, 2018, p. 101-102. Indica por eso Piñar cómo la identidad de la persona se configura hoy en la sociedad digital en torno al tratamiento de datos personales; ver Piñar Mañas, cit. p. 109. De hecho, ya Floridi concebía en 2011 la identidad personal como trasunto de la información; ver L. Floridi, «The informational nature of personal identity», *Minds and Machines* 21(4).

[45] J. Andreu Pérez, F. Deligianni, D. Ravi, G.Z. Yang, *Artificial Intelligence and Robotics,* 2017, p. 39, https://www.researchgate.net/publication/318858866_Artificial_Intelligence_and_Robotics

[46] Similar es el caso de la red LinkedIn, que tendía a mostrar más a hombres que a mujeres los empleos de mayor retribución.

«COMPAS», empleado por la justicia criminal de algunos Estados también norteamericanos, que está probado recomienda a los jueces un trato penal más severo respecto de encausados de raza negra[47]. O los de la Administración del Estado norteamericano de Indiana, detectada por la profesora de Harvard Virginia Eubanks, cuyos algoritmos negaban beneficios sociales a personas especialmente desfavorecidas[48].

Aunque es cierto que el sesgo puede asimismo resultar de intenciones abiertamente maliciosas[49], cual sería el caso citado de Indiana, si se llegase a descubrir que ese algoritmo se hubiera diseñado de intento para negar ayudas a los más pobres (es lo que de hecho sospecha Eubanks en su libro).

A los factores de sesgo indicados, y ya en un trabajo clásico de 1996, el informático Batya Friedman y la filósofa Helen Nissenbaum añadían uno más: aspectos derivados de un determinado «contexto de uso», que a su vez derivan de avances tecnológicos o bien de cambios introducidos respecto de los usuarios o beneficiarios de los sistemas en cuestión[50]. Un buen ejemplo de este último supuesto es el del algoritmo que va subiendo el precio de nuestro billete de avión en función de las sucesivas búsquedas que hacemos, desde la misma dirección de Internet, al detectar nuestra avidez por comprarlo.

Sea por unas o por otras causas, lo relevante aquí es que el sesgo algorítmico puede terminar generando discriminación, es decir, un trato injustificadamente más favorable para determinados grupos o personas[51].

[47] Las citas al respecto son abundantísimas. Por todas, ver R. Caplan, J. Donovan, L. Hanson, J. Matthews, *Algorithmic Accountability: A Primer*, 2018, p. 10, https://datasociety.net/library/algorithmic-accountability-a-primer/

[48] V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press, 2018.

En línea semejante se sitúa la noción de *technological redlining* (o «marcado tecnológico»), que conduce a resultados inicuos, a partir de discriminaciones contra negros, hispanos o nativo-americanos: ver S. U. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*, NYU Press, 2018.

[49] M. Brundage et al., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, 2018, p. 5, https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf

[50] B. Friedman, H. Nissenbaum, *Bias in Computer Systems*, 1996, p. 7, https://nyuscholars.nyu.edu/en/publications/bias-in-computer-systems

En la misma línea, EU Commission-Joint Reserch Centre, cit. p. 58.

[51] Un relevante informe oficial del Gobierno italiano lo recalca indicando cómo el sesgo puede generar «una distribución no homogénea de oportunidades»; ver AGID, *Libro Bianco sull´Intelligenza Artificiale al servizio del cittadino*, 2018, p. 64, https://ia.italia.it/assets/librobianco.pdf

La privacidad, por supuesto la privacidad, es otro de los derechos más frontalmente amenazados por los avances en inteligencia artificial.

Una muestra bien simple y conocida es la de las grabaciones «por defecto» de las voces de «sus amos» que realizan dispositivos inteligentes del tipo de Alexa de Amazon[52]. La grabación de voz es singularmente intrusiva. El uso ulterior de esos archivos sonoros puede comprometer gravemente a cualquiera, máxime cuando su obtención tiene en muchas ocasiones lugar en las esferas más propias de nuestra intimidad, como es el mismo hogar. No estamos además aquí ante un asalto si se quiere «lógico» o «digital» a nuestra privacidad, en cuanto tal limitado a los archivos que almacenemos en nuestro móvil u ordenador; el asalto ahora ya es *físico*, analógico, tangible, plasmado en nuestra vida real y no circunscrito a ese mero reflejo de la misma que es nuestra proyección digital. El cambio es bien relevante; tanto, que la Agencia de datos británica ha subrayado cómo la inteligencia artificial ha venido a «cambiarle el paso a [la privacidad y] la protección de datos»[53].

Con particular clarividencia, la mencionada Agencia británica especifica estas amenazas de la inteligencia artificial para la protección de datos[54]:

Son ya abundantísimos los trabajos que exploran esta cuestión. Entre los más influyentes, pueden citarse: T. Calders, I. Žliobaitė, «Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures», *Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases* 43, 55–56, 2013; D. Citron, F. Pasquale, «The Scored Society: Due Process for Automated Predictions», *89 Washington Law Review* 1, 8–18, 2014; F. Pasquale, *The Black Box Society: The Secret Algorithms that Control Money and Information*, Harvard University Press, 2015; S. Barocas, A.D. Selbst, «Big Data's Disparate Impact», *104 California Law Review 671 (2016)*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899

Como la OCDE señala: «Para muchos sistemas de inteligencia artificial, un entrenamiento con mayor cantidad de datos puede mejorar la precisión de sus resultados y reducir así el riesgo de discriminación por muestras sesgadas»; no obstante, sigue diciendo la OCDE, y como aquí señalamos a propósito de la privacidad: «Cuantos más datos se recojan, mayores serán los riesgos para la privacidad de los sujetos afectados». Ver *Artificial Intelligence in Society*, 2019, https://www.oecd-ilibrary.org/sites/eedfee77-en/1/2/4/index.html?itemId=/content/publication/eedfee77-en&mimeType=text/html&_csp_=5c39a73676a331d76fa56f36ff0d4aca&itemIGO=oecd&itemContentType=book

[52] Entre otras muchas posibles fuentes, ver https://www.independent.co.uk/life-style/gadgets-and-tech/news/amazon-alexa-echo-listening-spy-security-a8865056.html

[53] ICO, cit. p. 9 y siguientes.

[54] ICO, cit., p. 9 y siguientes.

En línea semejante se ha manifestado la AEPD, en su ya citada Guía en la materia. La Guía aborda los aspectos más relevantes en la relación Inteligencia artificial-Protección de datos que deben ser tenidos en cuenta desde el diseño y presta especial atención, entre otras, a las facetas más claramente afectadas por algoritmos «inteligentemente

• Primero, el mismo uso de algoritmos, como ya hemos visto, genera lo que algunos han llamado «impredecibilidad desde el diseño», el hecho de que los resultados diverjan de lo que los diseñadores hubieran previsto al introducir los insumos. Por ejemplo, un sesgo. Esto entra en conflicto con un principio clave de la protección de datos, como es la finalidad del tratamiento: ¿Cómo justificar que un ente público recoja datos para otorgar por ejemplo una ayuda social y que, a resultas del proceso algorítmico, el resultado sea denegarla?

• Segundo, nuestra bien conocida «Black box» o caja negra. ¿Cómo cohonestar el «misterioso» proceso algorítmico con otra exigencia basilar de la privacidad cual es la transparencia en el tratamiento de la información personal?

• Además, y tercero, el algoritmo se caracteriza por recoger toda la información posible (incluya o no datos personales) y que procede, bien del propio interesado, bien de observaciones o inferencias efectuadas por el sistema en sí. Esta «voracidad algorítmica» casa mal con otro postulado básico de la privacidad, la llamada minimización del dato, que justamente busca lo contrario. Y también con la calidad del dato, pues difícilmente todo ese océano digital será adecuado a los fines para los que se está tratando.

• Incluso topa otra vez con la transparencia, pues el interesado podría ignorar por completo que un sistema inteligente está observando su conducta, en línea o incluso física, como hemos visto a propósito de

débiles» pero enormemente voraces en datos y que adoptan decisiones en una «caja oscura» hermética al exterior (potencial fuente a su vez de sesgos de cualquier índole y subsiguientes discriminaciones). Dichas facetas son: la legitimación para el tratamiento, el ejercicio de derechos, la exactitud, la minimización de datos, la evaluación de impacto y el análisis de la proporcionalidad del tratamiento. Es de mucho interés el análisis sobre el ciclo de vida del sistema artificialmente inteligente, que puede llevar a que en una u otra fase esté o no sujeto a la normativa de protección de datos; también el reparto de roles entre desarrolladores, validadores, explotadores, etc., a fin de definir su correspondiente naturaleza de responsables o encargados y que concluye con la imposibilidad de trasladar la responsabilidad al propio sistema de inteligencia artificial (lo que permite inferir, correctamente a mi juicio, como en su momento se analiza en este trabajo, que la Agencia niega virtualidad a alguna suerte de «personalidad electrónica» a estos efectos). Ver AEPD, cit., https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf

Diversos autores vienen proponiendo la especificación de alguna de esta técnicas, en concreto la evaluación de impacto en protección de datos, en relación con los algoritmos inteligentes, mediante lo que denominan «evaluación de impacto algorítmico». Para una buena síntesis, comprensiva de algunas innovadoras propuestas, ver M. E. Kaminski, G. Malgieri, «Algorithmic Impact Assessments under the GDPR: Producing Multi-layeredExplanations», https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456224

Alexa. Y si desconoce todo esto, es claro que el interesado no podrá entonces ejercer ninguno de sus derechos al respecto.

Otro ejemplo más en este mismo sentido de observación «física», el terrible de la ciudad china de Shenzen, donde una de las tecnologías artificialmente inteligentes más potencialmente intrusivas para la privacidad, el reconocimiento facial, se está utilizando ya para identificar, multar y notificar por mensajería instantánea a peatones que cruzan las calzadas por lugares indebidos[55].

O lo que es, al menos, tan serio, el interesado podría ignorar también que el sistema inteligente está realizando inferencias sobre esa conducta digital o física. Es lo que se llama trazado de perfiles, que la inteligencia artificial (especialmente en su combinación con las tecnologías Big Data) puede llevar a cabo gracias a su potencial para generar correlaciones entre las miríadas de piezas de información que maneja. Ya citábamos antes el caso del perfil que un banco puede trazar de nosotros para concedernos o denegarnos una hipoteca, pero podemos añadir otros,

[55] https://www.independent.co.uk/news/world/asia/china-police-facial-recognition-technology-ai-jaywalkers-fines-text-wechat-weibo-cctv-a8279531.html

La High Court de Inglaterra y Gales validaba en el caso *Bridges* [2019] EWHC 2341 (Admin), sentencia de 4 de septiembre de 2019, el uso en pruebas por la policía galesa de cámaras fijas y móviles de reconocimiento facial sitas en lugares públicos, a modo de control policial y contra listas predeterminadas de personas en busca y captura o sospechosas de ciertos delitos; las razones fueron su uso por tiempo limitado, para fines específicos y también limitados, así como que se procede al borrado inmediato de los datos de quienes no figuran en la lista predeterminada en cuestión.

En cambio, la primera sentencia francesa sobre el empleo de este técnica, procedente del Tribunal administrativo de Marsella (nº 1901249, 27 de febrero de 2020), declaraba contrario a la privacidad un sistema de control de accesos a dos institutos de educación secundaria implantados por la Región de Provenza-Alpes-Costa Azul. Aun cuando el sistema se basaba en el consentimiento de los alumnos y, en su caso, de sus representantes legales, no podía considerarse válido a la vista de la autoridad ejercida por los responsables de ambos centros educativos. Además, y sobre todo, se trataba de un sistema irrespetuoso con el principio de proporcionalidad establecido en la normativa europea de protección de datos (Reglamento (UE) 2016/679, de 27 de abril de 2016, General de Protección de Datos, RGPD, art. 9), por cuanto la Región fue incapaz de demostrar que un sistema de identificación tradicional, quizá complementado con videovigilancia, no sería suficiente para conseguir el objetivo pretendido. Ver https://forum.technopolice.fr/assets/uploads/files/1582802422930-1090394890_1901249.pdf

Mucho más controvertida es una variedad del reconocimiento facial, el llamado reconocimiento de emociones (*affect recognition*), que, como algunas fuentes denuncian, no solo puede generar dosis acentuadas de discriminación, sino que carecería de base científica acreditada que justifique su uso (pese a ello creciente en países como los EE.UU.). Ver *AI Now Report 2019*, p. 12, https://ainowinstitute.org/AI_Now_2019_Report.pdf

como el que sobre si una mujer estaba o no embarazada elaboraba la cadena estadounidense de supermercados Target, a partir de los productos que sus clientas adquirían en sus tiendas.

• En quinto y último lugar, y esta vez es el Parlamento Europeo quien lo advierte, sujetos malintencionados pueden tratar de quebrantar la confidencialidad de un sistema inteligente, tratando de obtener información sobre los datos de entrenamiento del algoritmo o de extraer el modelo mismo del sistema. Unos y otro pueden obviamente contener datos personales, con lo que otro principio más de la privacidad, en este caso la seguridad, quedaría comprometido[56].

• Tras lo expuesto, nos adentramos en los desafíos que la inteligencia artificial nos plantea a todos en nuestra condición de consumidores y usuarios, y que la Organización europea de consumidores, cifra, además de por supuesto en los ya citados de discriminación o privacidad, en los siguientes[57]:

• Asimetrías de poder, que sitúan al consumidor en situación de clara desventaja, a la vista del plus de información que una plataforma puede llegar a adquirir gracias al algoritmo inteligente o incluso de la falta de habilidades digitales de muchos consumidores, sobre todo los especialmente vulnerables.

• «Confinamiento digital» de grupos más o menos amplios de consumidores en las plataformas de proveedores normalmente poderosos, quienes gracias a sistemas inteligentes pueden así impedir la movilidad del consumidor hacia otros proveedores donde puedan obtener mayor calidad o precio; práctica que lógicamente implica también una posible merma de la competencia en detrimento de empresas menos avanzadas tecnológicamente, que tendrán dificultades para detectar estas prácticas y reaccionar frente a ellas.

• El ya bien conocido «ajuste algorítmico de precios», que, como mínimo, provoca que el consumidor no pueda utilizar Internet para obtener el mejor precio posible, pues el sistema en cuestión siempre tenderá a igualarlos por arriba. Y, en el peor de los casos, cuando este ajuste es concertado, puede incluso responder a un arreglo colusorio, evidentemente contrario a las reglas de competencia.

[56] Parlamento Europeo, *Understanding algorithmic decision-making*, 2019, p. III y 55, https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2019)624261

[57] BEUC, *Automated Decision Making and Artificial Intelligence. A Consumer Perspective,* 2018, p. 5-10, https://www.beuc.eu/publications/beuc-x-2018-058_automated_decision_making_and_artificial_intelligence.pdf

• La Organización europea de consumidores alude por último a que un sistema algorítmico puede también generar «discriminación por precio», pues gracias a la posibilidad de rastrear precios en línea y de cruzarlos con los datos de navegación del consumidor, el sistema estará en condiciones de determinar con exactitud el precio que dicho consumidor podrá o querrá pagar.

En abierta discrepancia con este punto de vista, el norteamericano Center for Data Innovation asegura que la discriminación por precio puede llegar a ser beneficiosa para el consumidor, ya que también puede operar a la baja. Así, merced a la información adicional que el algoritmo proporciona, las plataformas tenderán a cobrar precios mayores a los consumidores cuya demanda es inelástica, pero estarían por el contrario incentivadas a bajarlos respecto de aquellos consumidores más sensibles al precio, que son por cierto los de menores ingresos[58].

Debemos sin embargo preguntarnos si está suficientemente justificado que ese beneficio para los consumidores de menores ingresos deba operar a costa de que, para todos los demás, el precio siempre sea el que más favorezca a la plataforma, es decir, que, en todos esos otros casos, la plataforma obtenga un plus de beneficio.

El último de los retos de la inteligencia artificial que comentaremos es el referido a la responsabilidad que deba derivar de los daños que un determinado sistema pueda causar.

En este sentido, no revisten especial dificultad los daños de naturaleza criminal. Hoy por hoy, y hasta un futuro que, como hemos visto, está aún lejano, ningún sistema inteligente puede equipararse a una *mens criminis* humana, es decir, a un cerebro que, como el humano, acumule el conocimiento y voluntad propios del dolo penal. Por consiguiente, siempre habrá un «hombre de atrás» (desarrollador de software, fabricante o usuario) a quien responsabilizar de los daños de esta naturaleza que se puedan haber causado, como por ejemplo un asesinato perpetrado a través de un dron[59].

¿Y en el caso del delito cometido por imprudencia? Cierto que tampoco el sistema inteligente puede ser imputado por gestionar con grave negligencia un riesgo, por ejemplo una ciclista de 49 años que atravesaba un paso de cebra de Tempe (Arizona, EE.UU.), como sucedió al vehí-

[58] Center for Data Innovation, *Competition and Consumer Protection in the 21st Century Hearings*, Project Number P181201, 2019, p. 28, http://www2.datainnovation.org/2019-ftc-competition-consumer-protection.pdf

[59] En este mismo sentido, T. King et al., «Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions», 2018, p. 23-25, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3183238

culo autónomo de Uber Technologies que la atropelló en marzo de 2018 y le causó la muerte. También aquí puede identificarse al diseñador, fabricante o usuario que hubiera debido prever lo que sucedió. La diferencia es que esta vez el «hombre de atrás», al no operar con dolo penal, podría tratar de ampararse en la imprevisibilidad del resultado, servida por la ya conocida impredecibilidad del algoritmo, la muy citada Black box. Un argumento que se irá viendo potenciado conforme vaya aumentando la autonomía de los sistemas inteligentes frente a los agentes humanos que los creen o utilicen[60]. Y un argumento que, al suscitarse a propósito de la culpa o negligencia, puede también esgrimirse respecto de la responsabilidad de índole estrictamente civil[61].

Responsabilidad civil, en concreto por pérdidas pecuniarias dimanantes de decisiones de inversión recomendadas por un sistema inteligente, es justamente lo que se dilucida en el primer caso judicial mundial de esta índole, que enfrenta a Li Kin-kan, magnate hongkonés, y la empresa Tyndaris Investments, en un juzgado mercantil de Londres. Según alega el demandante, el sistema, al que se encomendó la gestión de 2.500 millones de dólares, y que había de doblar esa inversión tras un tiempo, comenzó a perder dinero muy poco después de empezar a funcionar para Kin-kan, a fines de 2017, incluidos 20 millones de dólares en un solo día, el 14 de febrero de 2018[62]. La demandada alega que esas promesas de ganancia jamás se hicieron. La sentencia que en su momento se dicte comenzará a arrojar luz sobre el impacto de los sistemas inteligentes en los escenarios clásicos de la responsabilidad civil. Un impacto mucho más complejo que el antes expuesto del ámbito penal[63].

[60] P. De Filippi, A. Wright, *Blockchain and the Law. The Rule of Code*, 2018, p. 8.

En el caso citado de Arizona, la fiscal encargada del caso decidió no presentar cargos contra Uber Technologies, aunque no por los argumentos mencionados, sino por falta de suficientes pruebas, lo que no es óbice para que ordenara su obtención, para ulterior investigación en instancias superiores, y por si incluso pudiera resultar responsable criminalmente el conductor «de reserva» que viajaba en el vehículo autónomo. Ver https://assets.documentcloud.org/documents/5759641/UberCrashYavapaiRuling03052019.pdf

[61] Así lo hace el estudio promovido desde la Comisión Europea y elaborado por el Expert Group on Liability and New Technologies – New Technologies Formation, *Liability for Artificial Intelligence and Other Emerging Technologies*, 2019, p. 32-34, https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608

[62] https://www.bloomberg.com/news/articles/2019-05-06/who-to-sue-when-a-robot-loses-your-fortune

[63] Puede encontrarse un completo y bien documentado análisis de los distintos escenarios de responsabilidad civil, desde la específica perspectiva de tecnologías como la

El primero de esos escenarios, bien es sabido, es el de la posible responsabilidad objetiva del fabricante. Sin problema si estamos ante un producto y resulta ser defectuoso, pues así se desprende de las directivas europeas en vigor, por ejemplo la de maquinaria (Directiva 2006/42/CE de 17 de mayo de 2006) o la de seguridad general de los productos (Directiva 2001/95/CE de 3 de diciembre de 2001). Ahora bien, como es bien sabido, no existe en esas normas criterio específico alguno de seguridad respecto de productos con software encastrado, que funcione mediante algoritmos inteligentes, siendo además bien claro que estas normas sobre responsabilidad no se extienden en absoluto a servicios digitales defectuosos[64].

Ante la ausencia de defectos, la siguiente opción es el riesgo que el sistema en sí pueda o no causar. De nuevo sin problema, por ejemplo, respecto de los vehículos inteligentes, pues la conducción de automóviles es una de las actividades paradigmáticas de generación de riesgo para personas y bienes. El criterio del riesgo resulta sin embargo inútil para usos en principio inocuos de un sistema inteligente, como puede ser el de, pongamos por caso, recomendar el consumo de una determinada marca de agua mineral.

Si en estos últimos supuestos, y por la razón que fuera, llegara a generarse algún daño, porque, siguiendo con el ejemplo, el manantial del agua en cuestión resultase estar contaminado, no restaría sino volver al ya expuesto criterio de la culpa o negligencia. Aunque también aquí sería necesario que el daño hubiera sido previsible. Y, como en el ámbito criminal, igualmente en el civil la previsibilidad puede fácilmente topar con la oscura impredecibilidad del algoritmo.

Llegados a este punto, es momento de concluir que el Derecho, al menos desde los ángulos que hemos ido examinando, está ya sometido a un considerable zarandeo por parte de la inteligencia artificial. Y también de preguntarnos qué hacer al respecto.

III. ¿Puede en realidad hacerse algo desde el Derecho?

Hasta los foros más escépticos con la regulación de la inteligencia artificial, no solo piensan que es posible actuar al respecto, sino que exigen que así sea[65].

Internet de las cosas y la propia inteligencia artificial en P. Llaneza González, *Seguridad y responsabilidad en internet de las cosas (IoT),* Bosch, 2018.

[64] BEUC, cit., p. 16-17.

[65] Por todos, basta mencionar el caso del estadounidense Center for Data Innovation, varias veces citado en este trabajo.

De hecho, comienzan ya a abundar soluciones de índole tecnológica, en forma de mecanismos matemáticos o de aplicativos para la industria[66]. Y asimismo fórmulas organizativas como identificación de contextos problemáticos, introducción de pruebas antisesgo o fomento de la inversión en estas áreas[67].

No obstante, la experiencia de ya más de veinte años de relación intensa, en cuanto que socialmente generalizada, entre tecnologías digitales y Derecho, nos ha enseñado una importante lección[68]. Es la de que, como cualquier campo de acción humana, la norma jurídica tendrá que entrar a regular también determinados aspectos de la inteligencia artificial.

Y esa normativa adoptará, bien la forma de *soft-law*, o autorregulación por parte de la propia industria, a través de estándares y normalizaciones tipo ISO y similares[69] o de códigos de buenas prácticas[70]; bien la de normas jurídicas «tradicionales», es decir, con respaldo estatal, pues, como hemos visto, los principios y derechos en juego (dignidad, igualdad, privacidad…) son al fin y al cabo de la mayor importancia[71].

Si a la vista de la ya profusa literatura sobre regulación jurídica de la inteligencia artificial, tuviéramos que destacar dos principios por encima de todos los demás que se viene proponiendo, no cabe duda de que serían el de *centralidad de la persona humana*, en forma de *control* sobre los sistemas inteligentes (como por ejemplo propugna en Japón

[66] M. Whittaker et al., *AI Now Report 2018*, p. 24-27, https://ainowinstitute.org/AI_Now_2018_Report.pdf

[67] McKinsey Global Institute, *Notes from the AI frontier: Tackling bias in AI (and in humans),* 2019, p. 6, https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Tackling%20bias%20in%20artificial%20intelligence%20and%20in%20humans/MGI-Tackling-bias-in-AI-June-2019.ashx

Ambos tipos de medidas pueden encontrarse igualmente en ICO, cit., p. 86 y ss.

[68] R. Calo subraya esta obligada reflexión en «Robotics and the Lessons of Cyberlaw», 2015, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2402972

Existe una versión española de este trabajo: R. Calo, «La robótica y las lecciones del Derecho cibernético», *Revista de privacidad y Derecho digital*, nº 2, enero de 2016.

[69] A. Bertolini, E. Palmerini, «Regulating Robotics. A Challenge for Europe», EU Parliament, Directorate General for Internal Policies, *Workshop on Upcoming Issues of EU Law*, 2014, p. 194 y siguientes. http://www.robolaw.eu/

[70] *Algorithmic Accountability*…, cit., p. 24-25

[71] *Algorithmic Accountability*…, cit., p. 24-25

M. Guihot et al., cit. sugieren acciones en esta misma línea, si bien, en especial respecto de las llamadas Big Techs, y en atención al acentuado desequilibrio de información que éstas acumulan en su favor, propugnan soluciones más basadas en la mera sugerencia (*nudging*) que en medidas abiertamente vinculantes.

la iniciativa Human Centric AI[72] o entre nosotros el Consejo de Europa[73]); y el principio de *responsabilidad proactiva* (o *accountability*) por parte de dichos sistemas, según preconiza entre muchas otras fuentes el propio Parlamento Europeo[74]. Dos principios, debo advertir, que, como los demás que seguidamente referiremos, constan también en las declaraciones de naturaleza ética que más atrás mencionábamos, lo que en nada desmerece, por cierto, su paralela virtualidad en el mundo del Derecho.

Control humano y responsabilidad proactiva no son valores situados en un mismo plano jerárquico. La centralidad humana (y el consiguiente control por su parte de la inteligencia artificial) opera como una suerte de supraprincipio, en tanto y en cuanto la responsabilidad proactiva de los sistemas inteligentes justamente ha de orientarse en aras de tales centralidad y control. De nada serviría una «accountability» subordinada al abuso del poder estatal o exclusivamente al beneficio de la empresa privada. Es además este particular engranaje entre centralidad humana y «accountability» el que contribuye a generar otro valor fundamental en este ámbito, cual es el de *confianza*: ¿cómo pretender, por el contrario, que las personas nos fiemos de sistemas inteligentes que pudieran actuar *legibus solutus*, al margen de toda justificación y de toda consecuencia por las decisiones que lleguen a adoptar?

¿Qué debemos entender, por cierto, por «accountability» o responsabilidad proactiva de un sistema inteligente? Esa misma fuente del Parlamento Europeo la concreta en «la obligación de justificar sus decisiones y la posibilidad de enfrentarse a sanciones si tal justificación resultase inadecuada»[75]. Un concepto bien parejo, la verdad, a la clásica obliga-

[72] Gobierno de Japón, *Social Principles of Human Centric AI,* https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/humancentricai.pdf

[73] Consejo de Europa, *Addressing the Impacts of Algorithms on Human Rights*, Draft Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, 2018, https://rm.coe.int/draft-recommendation-of-the-committee-of-ministers-to-states-on-the-hu/168095eecf

[74] Parlamento Europeo, *Understanding algorithmic…,* cit., p. III.

Entre esas «muchas otras fuentes» referidas, baste citar: F. Doshi-Velez, M. Kortz et al., «Accountability of AI Under the Law: The Role of Explanation», 2019, https://arxiv.org/abs/1711.01134; R. Caplan et al., *Algorithmic Accountability*…, cit.; M. Whittaker et al., *AI Now Report 2018*, cit.; o Center for Data Innovation (CDI), *How Policymakers…,* cit..

[75] En la misma línea cabe mencionar las definiciones de la World Wide Web Foundation o el centro de investigación Data & Society, si bien estas dos entidades lo hacen en este sentido equivalente a «responsabilidad», en cuanto necesidad de reparar daños. Ver World

ción de motivar sus actos, que el Derecho administrativo impone a los poderes públicos.

El principio de responsabilidad proactiva ha logrado la primacía sobre otros tres principios adyacentes, que ya resultan absolutamente tópicos en esa ya profusa literatura sobre nuestro tema:

- Uno es el de *lealtad* («fairness») algorítmica, que las mejores definiciones equiparan a «ausencia de sesgos indeseados»[76].
- El segundo es el principio de *transparencia*, que de nuevo el Parlamento Europeo considera como «disponibilidad del código del sistema y de la documentación correspondiente»[77].
- Y el tercero, el llamado principio de *explicabilidad*. De la mano, entre otros, del Berkman Klein Center de Harvard University, lo entenderemos como «interpretabilidad o comprensibilidad de un sistema inteligente»[78].

Este último principio, la explicabilidad, es sin duda el más importante de los tres, en cuanto que digno de una enorme atención doctrinal y completamente genuino de la inteligencia artificial, a la vista de la bien conocida opacidad algorítmica.

De entre los ríos de tinta que se han vertido ya sobre este principio, dos ideas me parecen especialmente relevantes. Una, la de que, por mucho que el proceso de decisión algorítmico no sea humano y resulte altamente opaco, no parece razonable imponerle exigencias superiores a

Wide Web Foundation, «Algorithmic Accountability, 2017, http://webfoundation.org/docs/2017/07/Algorithms_Report_WF.pdf; R. Caplan et al., *Algorithmic Accountability*…, cit., p. 22.

[76] Parlamento Europeo, *Understanding algorithmic…*, cit., p. 30.

[77] Parlamento Europeo, *Understanding algorithmic…*, cit., p. 30.

[78] F. Doshi-Velez, M. Kortz et al., «Accountability of AI», cit., p. 2 y 3. Ver también A. Campolo et al., *AI Now Report 2017*, p. 26, https://ainowinstitute.org/AI_Now_2017_Report.pdf; y Parlamento Europeo, *Understanding algorithmic…*, cit., p. 30.

Se pueden seguir tres técnicas principales para llevar a la práctica el principio de explicabilidad: a) El enfoque de la «caja negra», que analiza el comportamiento del sistema por así decir sin `abrir la tapa', es decir, sin ningún conocimiento de su código. Las explicaciones se construyen a partir de observaciones de las relaciones entre los inputs y outputs del sistema. b) El enfoque de la «caja blanca»: a diferencia del enfoque de caja negra, este enfoque asume que el análisis del código del sistema es posible. Un ejemplo de los primeros trabajos en este sentido es el sistema Elvira para la explicación gráfica de las redes bayesianas. c) Y el enfoque constructivo: en contraste con los dos primeros enfoques, que asumen que el sistema ya existe, éste consiste en diseñar el sistema inteligente teniendo en cuenta los requisitos de explicabilidad («explicabilidad desde el diseño»). Ver al respecto Parlamento Europeo, *Understanding algorithmic,* cit., p. IV.

las que en general prevén hoy los ordenamientos jurídicos para regular cualquier otro proceso de adopción de decisiones[79].

La segunda idea es la de que no estamos ante una ciega y mecánica posibilidad de exigir en toda circunstancia una explicación. Como el mismo Berkman Klein Center de Harvard indica[80], dicha explicación solo tendrá sentido si: bien un tercero ha sufrido un daño que, además, sea resarcible (tal daño resarcible estaría por ejemplo ausente en el algoritmo de un comparador de precios); o si concurre algún interés en la explicación ante sospechas fundadas de error en el sistema (por insumos inadecuados, resultados inexplicables o desconfianza sobre las rectas intenciones del sistema).

Sea como fuere, el principio de explicabilidad presenta serios inconvenientes. Me parecen de especial importancia algunos de entre los que el mencionado Center for Data Innovation detalla: En primer lugar, y a la vista de la enorme complejidad de algunos sistemas, incluso tecnólogos altamente formados podrían ser incapaces de comprender nada tras acceder a su interior. Segundo, algunos de estos algoritmos podrían estar protegidos por derechos de autor, dada su condición de software, lo que legítimamente podría llevar a sus titulares a rehusar proporcionar toda explicación. En tercer lugar, esa explicabilidad podría terminar beneficiando a agentes maliciosos que accedieran a la misma. Y cuarto, y, por regla general, interpretabilidad y precisión están en relación inversa en inteligencia artificial, por lo que reforzar la primera podría menoscabar el inmenso potencial innovador de estas tecnologías[81]. A los inconvenientes mencionados, los muy autorizados expertos norteamericanos Annany & Crawford añaden algún otro, del que destacaremos el que llaman «transparencia resistente», o deliberada intención de camuflar los datos importantes con una avalancha adicional de información inútil[82].

Gracias a que, en contraste con el de explicabilidad, el principio de «accountability» opera por así decir «a cierta distancia» del código tecnológico, éste puede sortear todos y cada uno de los inconvenientes expuestos y que precisamente se refieren de lleno al código. Además, supera las

[79] Así lo recuerdan un influyente estudio de la Comisión Europea, *Algo:aware Raising Awareness on Algorithms*, 2018, p. 23, https://actuary.eu/wp-content/uploads/2019/02/AlgoAware-State-of-the-Art-Report.pdf; y el Berkman Klein Center de la Universidad de Harvard, en F. Doshi-Velez, M. Kortz et al., «Accountability of AI», cit., p. 12.

[80] F. Doshi-Velez, M. Kortz et al., «Accountability of AI», cit., p. 4-5 y 12.

[81] Center for Data Innovation (CDI), *How Policymakers…*, cit., p. 9-13.

[82] M. Ananny y K. Crawford, «Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability», 2016, *New Media & Society*, 1 –17, p. 7.

claras limitaciones del principio de lealtad, insoslayables, si tenemos en mente que, como bien hemos visto, todo sistema inteligente opera por definición con sesgos. Por otro lado, es general entender la transparencia como un principio subsidiario, en cuanto que instrumental, de la propia «accountability»: aquélla serviría para fortalecer ésta[83]. Mientras que el par de fuerzas «justificación-sanción» permite afrontar con eficacia nuestra necesidad clave: hacer de la persona el centro de la inteligencia artificial, sujetando ésta a nuestro control. Todo esto es lo que explica esa primacía de la responsabilidad proactiva sobre sus principios adyacentes: lealtad, transparencia y hasta la «muy atractiva» explicabilidad[84].

Todos estos principios están llamados a ir plasmándose en normas jurídicas. Y, cómo no, a partir de esas normas, en jurisprudencia. Hasta ahora, eso sí, y con la sola excepción de los vehículos autónomos (como

[83] Por todos, ver F. Doshi-Velez, M. Kortz et al., «Accountability of AI», cit.; y M. Ananny y K. Crawford, «Seeing without…», cit., p. 2

[84] No puede en consecuencia extrañar que los muy pocos países que han comenzado a regular la inteligencia artificial, o a plantearse hacerlo, centren justamente sus iniciativas en el principio de responsabilidad proactiva. Es el caso de dos proposiciones legislativas de idéntico texto y título, *Algorithmic Accountability Act*, que en abril de 2019 se introdujeron en el Senado y en la Cámara de Representantes de los EE.UU. https://www.congress.gov/bill/116th-congress/house-bill/2231/all-info

En lugar de pivotar sobre la «accountability», la *Executive Order on Maintaining American Leadership in Artificial Intelligence*, de 11 de febrero de 2019, gira en cambio en torno al principio de confianza, como garantía de una innovación en inteligencia artificial que a la vez respete «los derechos civiles, la privacidad y los valores estadounidenses» (sección 6). Ver https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/ En este marco normativo se mueven los 10 principios sobre IA elaborados por la Oficina de política científica y tecnológica de la Casa Blanca estadounidense, publicados a comienzos de enero de 2020 y asimismo presididos por el principio de «confianza pública»; estos principios deberán guiar la regulación que subsiguientemente las Agencias del Poder ejecutivo pretendan imponer al sector privado. Ver https://www.technologyreview.com/2020/01/07/130997/ai-regulatory-principles-us-white-house-american-ai-initiatve/

Es también el caso de la Comunicación de la Comisión Europea de 8 de abril de 2019, COM(2019) 168 final, *Inteligencia artificial para Europa*, que insta a avanzar en este campo con pleno respeto a «los valores y derechos fundamentales de la UE» (aunque igualmente insiste en la necesidad de una «human-centric AI»). También el Informe del Grupo de expertos de alto nivel sobre inteligencia artificial nombrado por la propia Comisión Europea, de 8 de abril de 2019, insta a construir una inteligencia artificial «confiable». En este mismo sentido debe señalarse la Declaración de los (a la sazón) 28 Estados de la UE sobre Cooperación en inteligencia artificial, firmada entre abril y julio de 2018 (si bien menciona al mismo nivel el principio de transparencia), y el italiano *Libro Bianco sull´Intelligenza artificiale*…, cit.. Sobre todo ello, ver https://www.loc.gov/law/help/artificial-intelligence/europe-asia.php

atestigua la siguiente figura), esto no ha comenzado a suceder más que en un concreto ámbito, el de la privacidad[85].

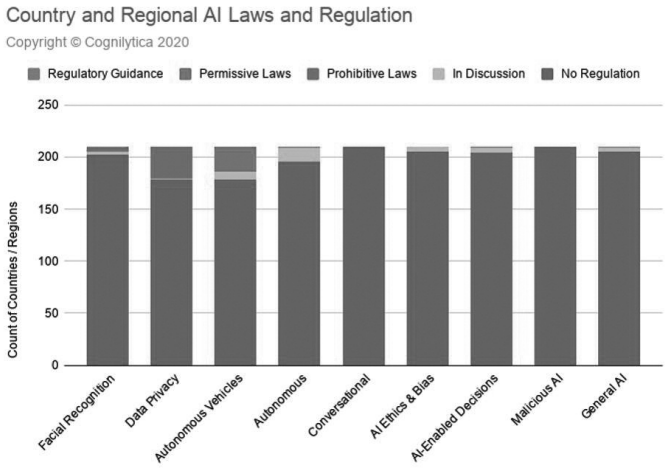

Figura 2. El panorama regulatorio mundial sobre IA.

Fuente: Worldwide AI Laws and Regulations 2020. https://www.cognilytica.com/2020/02/14/worldwide-ai-laws-and-regulations-2020/

[85] En este sentido cabe citar una pionera sentencia neerlandesa, procedente del Rechtbank Den Haag (La Haya, NL) y que data de 5 de febrero de 2020. En ella se declara contrario al derecho a la privacidad (art. 8.2 Carta Derechos Fundamentales UE, CDFUE) el algoritmo inteligente del llamado sistema «SyRI», empleado por las autoridades neerlandesas para predecir qué personas estarían en mayor riesgo de cometer fraudes en materia de viviendas públicas o de Seguridad social. El algoritmo recopilaba los perfiles de pasados infractores para lograr así «patrones de infractor». A partir de estos patrones, el sistema rastreaba bases de datos para identificar qué personas se acomodaban más a tales patrones predeterminados y poder así someterlas a un seguimiento más estrecho. El Tribunal considera todo ello contrario al derecho a la privacidad, como quiera que dicho seguimiento no obedecería a otras razones que la propia estigmatización derivada del algoritmo SyRI. Es la primera vez que un tribunal europeo ilegaliza el funcionamiento de un algoritmo inteligente a la luz de la normativa europea de derechos fundamentales. Cabe no obstante preguntarse si, aunque el Tribunal no menciona esta otra cuestión en absoluto, SyRI no resultaría asimismo contrario al principio de igualdad, también previsto por supuesto en CDFUE, en cuanto habría discriminado a los ciudadanos especialmente vigilados sobre la sola base de esa indicación algorítmica, lo que no parece justificación suficientemente objetiva ni razonable, máxime cuando resultaban generalmente ser personas económicamente desfavorecidas. Para la sentencia en *neerlandés*, ver https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:865&showbutton=true; para un buen resumen periodístico en *inglés*, ver https://www.dutchnews.nl/news/2020/02/governments-fraud-algorithm-syri-breaks-human-rights-privacy-law/

Todo lo cual nos aboca a comentar brevemente nuestra norma de cabecera en materia de privacidad, el RGPD de la UE. Además de la recién citada, ello responde a dos razones. La primera, que el RGPD constituye la primera norma jurídica del mundo en regular el impacto de la inteligencia artificial sobre la privacidad[86]. La segunda, que el RGPD bien puede considerarse el paradigma regulatorio de los dos principios clave que hemos venido analizando, el de centralidad o control por parte de la persona y el de responsabilidad proactiva, siendo a la vez ambos las dos columnas básicas sobre las que esta norma se asienta.

El principio de centralidad de la persona o de control respecto de sistemas inteligentes se instrumenta en el RGPD a través de dos vías. Una, el derecho que se concede a los afectados a no ser objeto de decisiones basadas únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que le afecten significativamente, con efectos jurídicos o no[87]. En otras palabras: un «derecho a una mínima intervención humana» en decisiones con impacto en la privacidad.

La segunda vía de control personal se concreta en el derecho de información a los interesados[88], que da pie a que fuentes muy autorizadas lo hayan llegado a catalogar como el «derecho a una explicación» frente al algoritmo[89]. Justamente por ello es apropiado afirmar que el RGPD consagra también, junto a los dos principios indicados, una exigencia de explicabilidad de los sistemas inteligentes, que a su vez concreta respecto de los mismos el general principio de transparencia que asimismo figura en su páginas. Este derecho obliga en esencia a los responsables de sistemas inteligentes que traten datos personales a proporcionar a sus usuarios una explicación acerca de la lógica seguida por aquéllos, así como de las consecuencias que para los usuarios puedan derivar de estas decisiones automatizadas. Es tan gráfica como ya muy popular la comparación de esta información con las tablas nutricionales que reglamentariamente figuran en las etiquetas de los alimentos[90].

[86] Países tan relevantes como Canadá comienzan a transitar también estas vías regulatorias: ver https://www.priv.gc.ca/en/about-the-opc/what-we-do/consultations/consultation-ai/pos_ai_202001/

[87] Art. 22 RGPD.

[88] Arts. 13 y 14 RGPD.

[89] Un muy buen ejemplo en M. E. Kaminski, «The Right to Explanation, Explained», 2019, https://scholar.law.colorado.edu/articles/1227/

[90] La idea se debe al experto en Marketing norteamericano B. Chudakov, ver L. Rainie, J. Anderson, «Code-Dependent: Pros and Cons of the Algorithm Age,» Pew Research Center, February 8, 2017, http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-thealgorithm-age/

A la vista de este derecho a la información algorítmica, comentaristas muy escépticos no solo huyen de celebrarlo, sino que predicen que las empresas que lo deban aplicar, bien reducirán la cantidad y complejidad de los datos y por ende la sofisticación del algoritmo para reducir costes, bien prescindirán del todo de emplear inteligencia artificial, con la consiguiente merma de competitividad europea; merma que surgiría muy especialmente, añaden, si los reguladores y los tribunales terminasen por interpretarlo en el sentido de exigir una explicación respecto de todas y cada una de las decisiones individualmente adoptadas por el algoritmo[91]. Con independencia del concreto sentido que esa interpretación acabe adquiriendo, es obvio que de la que lleven a cabo autoridades y jueces dependerá el que estos derechos de información adquieran verdadera fuerza o se queden en un brindis al sol[92].

En cuanto al principio de «accountability» o responsabilidad proactiva, se proyecta respecto de los sistemas inteligentes en el RGPD a través de otras dos vías. Primero, la necesidad de que estos sistemas incorporen pautas de privacidad desde el diseño y por defecto[93], es decir, que, como la creadora de estos conceptos, la canadiense Ann Cavoukian, manifestaba ya a mediados de los noventa, las normas sobre privacidad se «encastren» en el código de los sistemas y en los procesos de las organizaciones que los fabrican y emplean, de manera que la opción de reserva o «por defecto» que derive de su uso sea siempre la más respetuosa con la privacidad. Éste es el trasfondo de la bien actual corriente de situar todos los mecanismos sociales alrededor de la tecnología, desde el Derecho al Gobierno o a la Ética, entre otros, en el propio diseño de los sistemas, y que lleva a hablar de «Derecho desde el diseño», «Gobernanza desde el diseño» o «Ética desde el diseño». Una corriente que sin duda inspira por ejemplo al profesor norteamericano Lawrence Solum[94], cuando sugiere incorporar la inteligencia artificial al propio

La citada Guía AEPD de 2020 se refiere a ambas cuestiones (intervención humana e información) con profusión. Ver https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf

[91] Center for Data Innovation, *Competition and Consumer…, cit.,* p. 26-27.

[92] B. D. Mittelstadt et al., «The ethics of…», cit., p. 14. Varios de estos mimso autores exponen una visión aún más escéptica en S. Wachter et al., «Why a Right to Explanation of Automated Decision-Making Does not Exist», 2017, nota 6, p. 35, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469

[93] Art. 17 RGPD.

[94] L. B. Solum, «Artificially Intelligent Law», 2019, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3337696

diseño de las normas que deben regir un sistema inteligente, en forma de «Derecho artificialmente inteligente»[95].

La segunda vía de plasmación de la responsabilidad proactiva en el RGPD, a efectos de minimizar el impacto de la inteligencia artificial, es sin duda alguna la anonimización, es decir, el proceso tecnológico que, llevado a cabo sobre determinada información, permite «que el interesado no sea identificable, o [si lo era] deje de serlo»[96]. La gran atención que, por ejemplo, la Agencia británica de datos, dedica a este técnica en sus estudios sobre impacto en privacidad de la inteligencia artificial, es muestra más que suficiente de su utilidad e importancia en este campo[97].

Y, más allá de la privacidad, ¿qué sucede con otros campos del Derecho? Apuntaré dos ideas principales[98]. La primera se refiere al principio de igualdad, y es la de que los sistemas inteligentes no están desajustando las bases de la igualdad en su actual configuración, armada sobre los textos convencionales europeos, la Carta de la UE y las Constituciones

[95] Esta corriente anima también las propuestas de «inteligencia artificial desde el diseño» del Consejo de Europa y de algunos estudiosos del tema. Ver Consejo de Europa, *Declaration of the Committee…*, cit.; y R. Martínez, «Inteligencia artificial desde el diseño. Retos y estrategias para el cumplimiento normativo», *Revista Catalana de Dret Públic*, núm. 58, 2019.

[96] El modelo de anonimización más frecuentemente utilizado es la llamada *privacidad diferencial*, orientada a maximizar la privacidad pese a las deducciones que se puedan realizar sobre determinadas personas, minimizando al tiempo la pérdida de precisión de los datos, gracias a la introducción en las bases de datos de «ruido» matemáticamente modulado; ver Parlamento Europeo, *Understanding algorithmic…*, cit., p. 39.

El concepto tiene su origen en los trabajos de la tecnóloga Cynthia Dwork: «Differential privacy: a survey of results», *Proceedings of the 5th International Conference on Theory and Applications of Models of Computation (TAMC)*, Springer, 2008; y «Differential privacy», *Proceedings of the 3rd International Colloquium on Automata, Languages and Programming (ICALP)*, Springer, 2006.

[97] ICO, cit., p. 58-62.

[98] Aunque muy incipientemente, apunta ya también el impacto de la inteligencia artificial sobre el Derecho administrativo, como por ejemplo atestigua la francesa *Loi n° 2016-1321 de 7 octobre de 2016 pour une République numérique*. En armonía con las disposiciones semejantes de RGPD, esta Ley exige que los organismos públicos que adopten decisiones total o parcialmente basadas en sistemas algorítmicos, proporcionen a los ciudadanos medidas de transparencia como los criterios de tratamiento empleados y, si procediera, su correspondiente ponderación. La Comisión Europea abunda en estas mismas ideas, ver *Algo:aware…*, cit. p. 23.

Mientras que en el entorno del *Common law* han surgido ya muy relevantes aportaciones que instan a asegurar la responsabilidad, no solo de los entes públicos que emplean sistemas de IA en la adopción de decisiones públicas, sino de los propios agentes privados que los hubieran suministrado a dichos entes públicos. Ver K. Crawford, J. Schultz, «AI Systems as State Actors», *Columbia Law Review*, vol. 119, 2019.

estatales, además de en múltiples normas sectoriales. Ahora bien, y como muy bien anota el profesor de la Universidad de Amsterdam Frederik Borgesius, donde sí está impactando con fuerza el algoritmo inteligente es en la denominada discriminación indirecta, es decir, la no buscada de intento, conforme a su configuración por el Tribunal de Estrasburgo y el Derecho de la UE. La razón estriba en que, por una parte, esta modalidad de discriminación no responde a normas, sino a meros principios, con lo que resulta muy difícil de implementar en la práctica, máxime en entornos de opacidad algorítmica; una opacidad que justamente acentúa la segunda debilidad de la discriminación indirecta, cual es la necesidad de demostrarla: ¿cómo hacerlo en estos contextos, cuando ni siquiera se sabe que un algoritmo nos habría negado una beca, por, a modo de ejemplo, nuestra propia raza?[99] Otra cosa es que estas debilidades justifiquen algún tipo de reforma normativa al respecto, que quizá hiciera razonable una posible extensión a este concreto ámbito de la igualdad del ya expuesto derecho a una explicación previsto en la normativa sobre privacidad.

La segunda idea es la de que, merced a su severo impacto sobre tres áreas legales clásicas como son el consumo, la competencia y la responsabilidad civil, a su vez intrínsecamente imbricadas entre sí, el algoritmo inteligente está operando como vector de convergencia entre todas ellas. A esas tres áreas habría que añadir la protección de datos, aunque su análisis anterior nos releva ahora de más detalles. Ya se ha visto que ciertas prácticas algorítmicas con consumidores derivan fácilmente en prácticas anticompetitivas; la responsabilidad civil por un algoritmo defectuoso puede operar con absoluta generalidad respecto de consumidores; en tanto que autoridades de competencia como las alemana o francesa vienen últimamente demostrando que algunas de las empresas Big Tech pueden llegar a utilizar su «hegemonía sobre el dato» para menoscabar la competencia[100]. Por esto aciertan a mi juicio quienes reclaman una acción coordinada en este campo de los reguladores europeos sobre consumo, privacidad y competencia[101].

En lo que particularmente se refiere a la responsabilidad civil en inteligencia artificial, debemos recordar como dificultades la necesidad

[99] F. Z. Borgesius, *Discrimination, artificial intelligence, and algorithmic decision-making*, Consejo de Europa, 2018, p. 18-20, https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73

[100] https://www.ashurst.com/en/news-and-insights/legal-updates/algorithms-and-competition-law-franco-german-joint-study-published/

[101] BEUC, cit., p. 10.

de defecto en el escenario de responsabilidad objetiva (a su vez no aplicable hoy por hoy a sistemas con software encastrado ni a servicios digitales), así como la impredecibilidad algorítmica respecto de los escenarios de riesgo y culpa. Para superar estas dificultades, cabe detectar dos tendencias: una, la de agravar la responsabilidad, extendiendo el ámbito de la responsabilidad objetiva a supuestos de mero riesgo para la seguridad, incluso sin necesidad de defectos en el producto, todo ello al menos cuando el damnificado sea un consumidor[102]. Y otra, de sentido contrario, la de incluso aligerar los estándares de responsabilidad actuales, previendo sanciones más leves respecto de actuaciones que, aun habiendo generado un daño, estén fundamentalmente orientadas a la innovación y el beneficio de la sociedad[103].

A medio camino entre ambas tendencias se ha situado el Libro Blanco de la Comisión Europea sobre Inteligencia artificial, presentado en febrero de 2020, junto con todo un paquete de medidas en materia digital titulado *A Europe Fit for the Digital Age*[104]. El Libro Blanco se ancla en estos principios básicos, presididos por el evidente de aplicación general de cualesquiera normativas europeas o nacionales en vigor respecto de sistemas de inteligencia artificial (privacidad y PD, consumo, competencia, etc.):

• Regulación nueva y obligatoria para aplicaciones de inteligencia artificial de «alto riesgo», siendo definido éste por sectores y, acumulativamente, por los efectos –legales y daño físico o material– que puede llegar a generar en las personas[105]; así como con independencia de lo anterior, en función de circunstancias excepcionales, como pueden ser

[102] BEUC, cit., p. 17. En coherencia con ello, la propia Organización europea de consumidores preconiza la necesidad de considerar que los productos con contenidos digitales sean un «producto» a los efectos de la Directiva sobre responsabilidad de productos (Directiva 85/374/CEE, de 25 de julio de 1985), en tanto que ponerlos a la venta, por ejemplo en Internet, debiera a esos mismos efectos considerarse equivalente a ponerlos en circulación; ver BEUC, cit., p. 17.

[103] Center for Data Innovation, *Competition and Consumer…, cit.,* p. 29-30. Mientras que también en coherencia con su planteamiento recién expuesto, el Center for Data Innovation urge a que los reguladores centren su atención en los operadores, en cuanto agentes responsables de emplear e la práctica los algoritmos, en lugar de en los desarrolladores, puesto que son aquéllos quienes en el fondo adoptan las decisiones de mayor relieve acerca del impacto de los algoritmos en la sociedad; ver *Competition and Consumer…, cit.,* p. 30.

[104] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age_en

[105] En el mismo sentido, ver Expert Group on Liability and New Technologies – New Technologies Formation, *Liability for…,* cit., p. 40.

procesos selectivos de personal o identificación biométrica remota (en particular, reconocimiento facial)[106] y otras tecnologías de vigilancia. Esta regulación nueva exigirá asimismo actualizar la legislación europea en materia de seguridad y responsabilidad de productos y servicios[107].

- Aplicación de los principios y reglas de dicha regulación nueva y obligatoria a las distintas fases o facetas de despliegue de sistemas de inteligencia artificial, en concreto entrenamiento de algoritmos, almacenamiento de datos, transparencia informativa, robustez/precisión y supervisión humana.
- Etiquetado voluntario, o lo que es lo mismo, modelo similar a los escenarios de *soft-law* autorregulatorio existentes por ejemplo en el vigente RGPD, respecto de todas aquellas aplicaciones de inteligencia artificial que no conlleven alto riesgo.
- Y supervisión de la normativa que se adopte por parte de autoridades públicas (en coordinación con todas las actualmente existentes, pudiendo de hecho ser alguna de estas mismas).

En su conjunto, parece un esquema acertado, a la vista de la patente necesidad de aportar certeza y seguridad jurídicas que las ideas expuestas demuestran. Todo ello siempre y cuando, eso sí, la normativa que se termine por adoptar efectúe un deslinde especialmente cuidadoso de esa noción de «alto riesgo», so pena de debilitar seriamente el potencial de innovación de la UE en inteligencia artificial[108]. Otra cosa es tratar de superar los escollos para la responsabilidad acudiendo a un recurso que el propio Parlamento Europeo se ha llegado a plantear, sobre la base de

[106] En concreto sobre reconocimiento facial, se opta simplemente por promover un amplio debate a escala europea acerca de las ventajas e inconvenientes de su uso. Ver Comisión Europea, *Libro Blanco sobre Inteligencia Artificial*, 19 de febrero de 2020, https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

[107] En este mismo sentido, ver Expert Group on Liability and New Technologies – New Technologies Formation, *Liability for…*, cit., p. 34.

En lo que hace a la conveniencia de ajustar la regulación a las necesidades de sectores específicos, la iniciativa europea se alinea con la bien fundada e interesante recomendación de Y. S. Lee, B. C. Larsen, M. Webb, M. F. Cuéllar, *AI Regulation and Firm Behavior*, 14 de diciembre de 2019, https://voxeu.org/article/ai-regulation-and-firm-behaviour

[108] Comisión Europea, *Libro Blanco…*, cit.

No en vano, el quizá mayor valor añadido del mencionado trabajo de Y. S. Lee y otros radica en subrayar cómo la simple recepción de información acerca de los requisitos regulatorios vigentes o en proyecto en materia de inteligencia artificial disminuye la intención de las empresas de implantar inteligencia artificial en sus procesos de negocio. Ver Y. S. Lee, B. C. Larsen, M. Webb, M. F. Cuéllar, *AI Regulation,* cit.

un importante estudio de la Universidad de Pisa[109]. Me refiero a algo ya citado al comienzo de esta exposición: la posibilidad de atribuir una suerte de personalidad, una «personalidad electrónica», a determinados sistemas de inteligencia artificial, entre ellos robots suficientemente sofisticados.

Debo reconocer que la idea de una personalidad electrónica tiene el fuste necesario para ser tomada en serio. Lo demuestra la perfecta viabilidad de otras ficciones jurídicas, como la misma idea de persona jurídica, surgida hace mil años con el fin de superar la finitud temporal de la vida humana. Y ello por más que la persona jurídica no deje en el fondo de agrupar a personas humanas. Cierto igualmente que la idea de personalidad electrónica supondría traspasar una dogmática frontera, cual sería la de llegar a aplicarse a lo que en el fondo no dejan hoy de ser cosas[110]. Si bien es verdad que justamente vivimos hoy otros traspasos de fronteras dogmáticas, como la de que solo una persona física podía ser responsable penalmente: así lo atestigua el abandono del *societas delinquere non potest* para extender la responsabilidad criminal a la persona jurídica.

Además, la instrumentalidad de la personalidad electrónica sería la de perfeccionar la responsabilidad civil (e incluso criminal) respecto de los sistemas inteligentes, haciendo por ejemplo posible, que, como hoy sucede respecto de una sociedad anónima o una sociedad limitada, quien mantenga cualquier tipo de relación jurídica con un sistema de inteligencia artificial que merezca tal entidad, o sea objeto de un ilícito por su parte, pueda esperar un capital mínimo para hacer frente a la responsabilidad que la actuación de ese sistema pudiera generar.

Pese a todo lo dicho, y lo anticipaba casi al inicio, he llegado personalmente a la convicción de que esta idea de personalidad electrónica no

[109] Se trata del ya citado informe *Robolaw*, elaborado bajo auspicios de un proyecto de la Comisión Europea, que propugna el reconocimiento de una «personalidad electrónica», bajo determinadas condiciones, a robots. Y de la Resolución del Parlamento Europeo de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL).

La idea originaria no es sin embargo ni europea, ni tampoco reciente, pues ya la esbozó el más atrás citado profesor de Georgetown Lawrence B. Solum, nada menos que en 1992: ver «Legal Personhood for Artificial Intelligences», *North Carolina Law Review* 70, 1231.

[110] Pese a ello, no se debe olvidar que determinados sistemas inteligentes dotados de corporeidad y de una cierta sofisticación contrastan con la idea tradicional de «cosa» en ámbitos jurídicos, pues como Calo apunta, esos ingenios pueden llegar a poseer facultades cognitivas: piensan, aprenden, perciben, deciden; ver R. Calo, «La robótica y las lecciones…», cit., p. 155-157.

constituye un instrumento adecuado para resolver problemas como los expuestos[111].

Entiendo que es así porque este mecanismo extendería a sistemas o ingenios artificialmente inteligentes un atributo, la propia personalidad, que es patrimonio exclusivo de la esencia humana. Que, al hacerlo, diluiría por tanto esa exclusividad que los humanos tenemos sobre esa nuestra nota singular. Y que también por ello difuminaría el atributo capital de cuantos han de caracterizar un uso confiable de la inteligen-

[111] Esta opinión concuerda con la del grupo de expertos de alto nivel sobre IA promovido por la Comisión Europea, quienes consideran que la idea de personalidad electrónica socava los principios de centralidad humana y responsabilidad proactiva que deben guiar la Inteligencia artificial. Ver Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission, *Policy and Investment Recommendations for Trustworthy AI, 2019,* p. 41, https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence

Otros, como el British Standards Institute, ya se habían decantado en favor de exigir en todo caso ante estas situaciones la responsabilidad de personas humanas. Ya sea el fabricante, ya el distribuidor, ya el propietario, ya el usuario del sistema, etc. en función de las circunstancias que correspondan. El British Standards Institute expone esta tesis en su informe *Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems*, BS 8611:2016, abril de 2016. Es un documento de acceso restringido. Para una buena síntesis, ver https://www.theguardian.com/technology/2016/sep/18/official-guidance-robot-ethics-british-standards-institute

Ya desde el ángulo jurídico, y con el mismo resultado de denegar viabilidad a la idea de personalidad electrónica, se ha indicado que ésta aboca a «una especie de abuso de los ordenamientos jurídicos: mientras que las personas electrónicas disfrutarían de una gran cantidad de derechos (propios de los humanos) frente a las personas humanas, está por ver cómo las correspondientes obligaciones legales podrían invocarse en su contra»; ver J. J. Bryson, M. E. Diamantis, T. D. Grant, «Of, for, and by the people: the legal lacuna of synthetic persons», 2017, p. 285 y siguientes, https://link.springer.com/article/10.1007/s10506-017-9214-9

Una de estas autoras, J. J. Bryson, incide en concreto en cómo la personalidad electrónica sería un factor de desigualdad, al «blindar a las grandes empresas y a personas de alto poder adquisitivo frente a toda responsabilidad, en detrimento del ciudadano corriente»; ver J. J. Bryson, *The Past Decade, cit…*

En tanto que de nuevo J. J. Bryson afronta también el tema desde la perspectiva filosófica, al afirmar que hacer de los sistemas inteligentes «agentes (en lugar de pacientes) morales» constituiría, entre otras cosas por lo arriba señalado, «una opción menos ética» que la de mantener su actual estatuto de meras cosas; ver J. J. Bryson, «Patiency is not a virtue: the design of intelligent systems and systems of ethics», 2018, https://link.springer.com/article/10.1007/s10676-018-9448-6

El trabajo del Expert Group on Liability and New Technologies – New Technologies Formation, *Liability for…,* cit., p. 37-39 llega a la misma conclusión, que ancla en la innecesaridad de este mecanismo, pues la responsabilidad que se trata de asegurar puede hoy por hoy atribuirse siempre a algún agente humano.

cia artificial, como es la centralidad del ser humano. Y si esto sucede al hilo de la denominada inteligencia artificial débil, única con la que ahora convivimos, sobra decir cuánto más ocurriría así en el supuesto de llegar a coexistir con sistemas inteligentes fuertes, que, justo por serlo, fácilmente podrían aspirar a cotas incluso más dignificadas de «personalidad».

Hemos llegado al final. Para ello, hemos efectuado un recorrido a través de problemas y desafíos. Y a través de principios, que, cada vez más voces sostienen, deben ser el soporte que dote de fiabilidad a los sistemas de inteligencia artificial; por encima de todos, el principio de control y centralidad de la persona humana, para que sea la inteligencia artificial la que ponga en el centro a la persona, y no la persona quien deba vivir a expensas de la inteligencia artificial[112].

Inteligencia artificial, al fin y al cabo, es sin duda «conocimiento artificial», pero nunca será «sabiduría artificial». Como Sócrates nos enseñó, la sabiduría, la auténtica sabiduría, es una virtud, por eso la sabiduría es y será siempre un atributo exclusivamente humano. La máquina conoce, no sabe. La máquina desconoce la medida de su ignorancia.

[112] Un control que, como ha demostrado un interesante estudio empírico desarrollado por investigadores de las Universidades de Chicago y Pennsylvania, y aun cuando solo llegue a ser ligero, ayudará a reducir lo que estos mismos investigadores denominan «aversión» al algoritmo; ver B. J. Dietvorst, J. P. Simmons, C. Massey, «Overcoming Algorythmic Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them», 2016, p. 1, https://pubsonline.informs.org/doi/10.1287/mnsc.2016.2643

Un grupo de profesores de las Universidades de Bolonia, Florencia y Módena augura en este mismo sentido una Inteligencia artificial que «empodere a los consumidores», en lugar de solamente a agentes poderosos, como hasta ahora ha venido predominantemente sucediendo con las empresas, especialmente las Big Tech, como usuarias masivas de Inteligencia artificial; ver G. Contissa, F. Lagioia, M. Lippi, H. W. Micklit, P. Pałka, G. Sartor, P. Torroni, «Towards Consumer-Empowering Artificial Intelligence», https://www.ijcai.org/Proceedings/2018/714