



UNIVERSITAT DE
BARCELONA



Revista de Bioética y Derecho

Perspectivas Bioéticas

www.bioeticayderecho.ub.edu - ISSN 1886-5887

DOSSIER SOBRE INTELIGENCIA ARTIFICIAL, ROBÓTICA E INTERNET DE LAS COSAS

Las narrativas de la inteligencia artificial

The narratives of artificial intelligence

Les narratives de la intel·ligència artificial

MANUEL JESÚS LÓPEZ BARONI *

* Manuel Jesús López Baroni. Universidad Pablo de Olavide. Sevilla (España). E-mail: mjlopbar1@upo.es.

Copyright (c) 2019 Manuel Jesús López Baroni



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional.

Resumen

La expresión *Inteligencia Artificial* designa significados muy diferentes entre sí, cuando no antagónicos. En el presente artículo examinaremos algunos de estos sentidos, en concreto, Inteligencia Artificial como *trending topic*; como *big data*; como *sesgo*, como cuestión *sociolaboral*; como *ente sin conciencia*; como *ente con conciencia*; y, por último, como *disciplina convergente*. Cada una de estas expresiones nos muestra problemas de naturaleza muy diferente, algunos de los cuales intersectan con la bioética, de ahí el interés para nuestra disciplina. La normativa jurídica, dispersa y descoordinada, a veces incluso ausente, refleja las dificultades inherentes a no poder concretar exactamente a qué nos enfrentamos.

Palabras clave: inteligencia artificial; big data; conciencia; autonomía; vida artificial; sesgo.

Abstract

The expression Artificial Intelligence designates meanings that are very different from each other, if not antagonistic. In this article we will examine some of these meanings, in particular, Artificial Intelligence as “trending topic”; as “big data”; as a “bias”, as a “socio-labor issue”; as an “entity without conscience”; as an “entity with conscience”; and, finally, as a “convergent discipline”. Each of these expressions shows us problems of different nature, some of which intersect with bioethics, hence the interest for our discipline. Legal regulations, dispersed and uncoordinated, sometimes even absent, reflects the inherent difficulties of not being able to specify exactly what we are facing.

Keywords: artificial intelligence; big data; conscience; autonomy; artificial life; bias.

Resum

L'expressió Intel·ligència Artificial designa significats molt diferents entre si, quan no antagònics. En el present article examinarem alguns d'aquests sentits, en concret, Intel·ligència Artificial com *trending topic*; com *big data*; com a *biaix*, com a qüestió *sociolaboral*; com a *ens sense consciència*; com a *ens amb consciència*; i, finalment, com a *disciplina convergent*. Cadascuna d'aquestes expressions ens mostra problemes de naturalesa molt diferent, alguns dels quals intersequen amb la bioètica, d'aquí l'interès per a la nostra disciplina. La normativa jurídica, dispersa i descoordinada, a vegades fins i tot absent, reflecteix les dificultats inherents a no poder concretar exactament a què ens enfrontem.

Paraules clau: intel·ligència artificial; big data; consciència; autonomia; vida artificial; biaix.

1. Introducción

¿Por qué hemos de ocuparnos en bioética de la Inteligencia Artificial? En principio podríamos argumentar que afecta a los seres humanos, de ahí nuestro interés; sin embargo, también nos influye la inflación o el populismo, y no por ello suelen ser objeto de una reflexión específica en nuestra disciplina.

Una Inteligencia Artificial es básicamente, al menos por ahora, un programa informático, por muy sofisticado que este sea. Ahora bien, parece que en los últimos años los algoritmos que vertebran estos programas han dado un salto cualitativo.¹ Probablemente por ese motivo, en las facultades de informática recién se están planteando la necesidad de incorporar a sus planes de estudio cuestiones sobre ética, valores, etc., aunque descartando que la disciplina específica deba ser “bioética”. La causa de esta exclusión reside en que el prefijo “bio” implica, directa o indirectamente, que las entidades problemáticas deben estar *vivas*, como sucede con los seres humanos, los animales o los organismos modificados genéticamente, pero no con los robots.

Pues bien, intuimos que el motivo, aunque sea tácito o inconsciente, en todo caso no confesable públicamente, por el que la Inteligencia Artificial se comienza a estudiar *también* en bioética, como si la disciplina no fuese ya lo suficientemente amplia, es porque de alguna manera se presupone que estas entidades, sea cual sea su naturaleza, pueden alcanzar el estatuto de entidades vivas, con o sin conciencia, con o sin verdadera inteligencia, en algún momento de las próximas décadas.

Así, al igual que la bioética amplió progresivamente su campo de aplicación, biotecnología mediante, hasta abarcar la biología sintética o la edición genómica, la informática está entroncando en estos momentos con la bioética a través de la Inteligencia Artificial. De hecho, se podría sostener que ambas disciplinas, la biología sintética y la Inteligencia Artificial, tratan con entidades que están a medio camino entre lo vivo y lo inerte, entre lo programable y lo incontrolable, entre lo que se puede reproducir y lo que no, de ahí su imbricación con nuestra disciplina. Lo cierto es que, como si de un paradójico viraje se tratara, la bioética ha acabado fagocitando como campo de estudio la materia inerte *como si (o por si) cobrara vida*, analizándola, estudiándola y previendo múltiples escenarios que nos pueden afectar de forma estructural e irreversible.

En este sentido, debemos destacar cómo el Parlamento europeo, al proponer una Carta sobre Robótica, comenzó enumerando los principios de la bioética (autonomía, justicia,

¹ En este sentido, el gobierno francés ha destacado cómo en los últimos cinco años se ha producido una significativa aceleración en este campo, de ahí su propuesta de crear una cátedra jurídica sobre esta materia (AA.VV. 2017a).

beneficencia y no maleficencia) como reglas a seguir para los ingenieros en robótica (Parlamento europeo, 2017); o cómo la recopilación de reglas éticas que deben regir la Inteligencia Artificial se ha denominado “Principios de Asilomar para la Inteligencia Artificial” (AA.VV. 2017b), denominación de enorme simbolismo para nuestra disciplina, ya que en dicha localidad se acordó la primera moratoria científica de la historia a raíz de los experimentos con ADN recombinante, iniciando la integración de la biotecnología en la bioética.

En resumen, los paralelismos entre la bioética y la biotecnología, de un lado, y la bioética y la Inteligencia Artificial, de otro, son evidentes, de ahí la necesidad de un análisis interdisciplinar.

Pues bien, con objeto de aclarar a qué nos enfrentamos, vamos a analizar diferentes sentidos o significados de la Inteligencia Artificial. En concreto, estudiaremos la Inteligencia Artificial como *trending topic*; como *big data*; como *sesgo*; como cuestión *sociolaboral*; como *ente sin conciencia*; como *ente con conciencia*; y, por último, como *disciplina convergente*. Observaremos cómo el vocabulario, los problemas y las predicciones varían en función de qué sentido empleemos, de ahí que la incipiente normativa, esporádica, dispersa, y descoordinada, trate de hacer frente a esta avalancha de problemas, de naturaleza muy diferente entre sí, con más voluntad que fortuna.

2. Inteligencia artificial como *trending topic*

Periódicamente asistimos a un clímax de alarmismo como consecuencia de los avances en las nuevas tecnologías. Sucedió a finales de los noventa con la clonación, a principios de este milenio, con la nanotecnología, y ahora, con la edición genómica. Desde esta perspectiva, en la Inteligencia Artificial no sucede nada en especial, o al menos que no haya sucedido antes con la tecnología. Simplemente se sigue avanzando, progresando, refinando lo ya existente, etc., pero las especulaciones acerca de los riesgos para la humanidad son, en el mejor de los casos, una forma de pasatiempo académico. Podemos incluso concretar las fases por las que se ha transitado hasta alcanzar nuestro momento actual.

La primera fase comenzó cuando algunas de las figuras más mediáticas en ciencia o tecnología, como Stephen Hawkins o Elon Musk (fundador de Tesla), hicieron declaraciones más o menos alarmistas acerca de la inteligencia artificial. Sus opiniones reverberaron en los medios de comunicación y en las redes sociales, generando infinidad de comentarios, matices, posicionamientos a favor y en contra, etc., por lo que se creó un clima propicio para la acción (“Lejos de ser ciencia-ficción, la inteligencia artificial (IA) forma ya parte de nuestras vidas”, Comisión Europea, 2018b).

En una segunda fase, alguna institución relevante publica un informe analizando el estado de la cuestión, lo que obliga a otras entidades similares (universidades, Unión Europea, organismos de investigación, etc.) a seguir el paso para no perder relevancia pública ni prestigio (quien no tenga informe, no cuenta). El resultado es que en poco tiempo la opinión pública dispone de varios documentos con predicciones agoreras, lo que crea una desasosegante sensación de alarma que se contagia y expande. Es la fase espasmódica, algo parecido a una histeria colectiva, justo donde nos hallamos en estos momentos.

La tercera, inminente, sucede al agotarse las narrativas distópicas (ya no producen impacto, por lo que se trivializan o incluso ridiculizan), con lo que la temática decae progresivamente, hasta el completo olvido. Calificar como “inteligente” a todo aparato electrónico que pueda conectarse a internet, o que recopile información de sus usuarios, forma parte de este proceso de trivialización de las IA.

Una consecuencia nada inocente de la generación de este ambiente es que se logran más fondos para la investigación (cualquier político que se precie debe financiar estas tecnologías), e incluso algunos científicos, previa promesa de disponer del bálsamo de *fierabrás* para evitar escenarios apocalípticos, se ofrecen para aliviar nuestro estrés, prestigiándose ante la comunidad científico/política como chamanes singularmente capacitados para ahuyentar futuros terroríficos.

Al igual que sucede con otras modas, esta alcanzará su cénit dentro de poco y languidecerá en la noche de los tiempos, hasta que la próxima generación la recupere como entretenimiento cíclico y mediático.

3. Inteligencia artificial como *big data*

La Inteligencia Artificial suele aparecer junto a otra expresión, *big data*, con la que se alterna incluso en las titulaciones oficiales, como los másteres o los grados.

Este emparejamiento no es casualidad, ya que de alguna manera aclara el significado de la Inteligencia Artificial: se trata de recolectar y analizar ingentes cantidades de datos, desde los climáticos hasta los de consumo personalizado, pasando por el comportamiento de las bolsas, los insectos o el deporte, con objeto de poder hacer predicciones lo más fidedignas posibles. La supuesta “inteligencia” se basa en la presuposición de que, cuantos más datos se puedan procesar, más posibilidades habrá de anticiparse a una conducta, comportamiento o suceso. La apuesta por los superordenadores por parte de la Unión Europea (v. gr. creación de la Empresa Común de Informática de Alto Rendimiento), con su correspondiente neolengua (superordenadores a

prexaesca, exaesca y petaesca, etc.), así como la preparación para que “las arquitecturas informáticas clásicas se integren con los dispositivos informáticos cuánticos, por ejemplo, utilizando el ordenador cuántico como un acelerador de hilos de informática de alto rendimiento” (apt. 21), se puede situar en esta línea discursiva.²

El término clave aquí es el de “algoritmo”.³ Cuanta más complejidad algorítmica, más capacidad predictiva, de ahí que todos los esfuerzos se centren en correlacionar infinidad de variables para acertar en las previsiones de futuro. El sueño inconfesable es lograr el “demonio de Laplace”, esto es, una máquina que, previa recopilación de todas las variables posibles, sea capaz de predecir cualquier escenario de una forma completamente determinista.

Desde esta perspectiva, nos hallamos, en efecto, ante un hecho ciertamente singular, ya que nunca en la historia de la humanidad se ha podido recopilar tal volumen de información. Es previsible que la entrada en vigor del *internet de las cosas* (cualquier objeto será conectado a la red, desde la ropa hasta los vehículos, pasando por los frigoríficos o nuestros cuerpos), aumente el flujo de datos, y, por tanto, la *inteligencia*.

La preocupación, por lo que a la bioética se refiere, radica en que la información genética forma parte de este proceso de recolección y análisis de datos. Las bases de datos que dispongan de toda la información genética de una población podrán realizar predicciones que incidirán en aspectos nada triviales, como los contratos con las compañías de seguros (una prima personalizada se basará en la predisposición, o no, a determinadas enfermedades), la búsqueda de trabajo (los recursos humanos dispondrán de sofisticados software para relacionar el puesto a desempeñar con el perfil psicológico/genético de los candidatos); o incluso la búsqueda de pareja (programas capaces de hacer predicciones acerca de las enfermedades o habilidades de los potenciales hijos), etc. No es descartable que, en un futuro inmediato, un Estado cuente con la secuenciación del genoma de todos sus habitantes, con lo que el problema de los *big data*

² Reglamento 2018/1488/UE, de 28 de septiembre. Crea la Empresa Común de Informática de Alto Rendimiento Europea. DOL 8 octubre 2018, núm. 252.

³ En principio, un algoritmo es una unidad de información, un “programa”, por breve o simple que este sea, pero en realidad nadie sabe muy bien qué significa esta expresión, ya que cualquier conjunto de reglas, tanto para freír un huevo como para tomar decisiones automatizadas en la bolsa, sería propiamente un algoritmo. Con objeto de aclarar este problema, el FRA (2018), a la hora de definirlo, se ha centrado en el contexto de la informática, de ahí que haya entendido por algoritmo una secuencia de comandos en la que se introducen inputs y se generan outputs (el ejemplo que nos ofrecen es la introducción de una lista de personas, input, para que el ordenador las ordene en función de su edad, output).

trascenderá el mundo privado y se convertirá en un problema de naturaleza política, como sucede en estos momentos en China con el control informático de sus ciudadanos.

Estos potenciales escenarios explican que, en su momento, el Convenio europeo de bioética prohibiera emplear la secuenciación genética más allá de su utilidad en el campo de la medicina; o que el Reglamento de protección de datos personales de la Unión Europea muestre especial preocupación por la recolección de datos genéticos (v. gr., el art. 9.1).

Pero la capacidad predictivo/biológica de los Big Data no se detiene en la información genética, ya que puede alcanzar otras fuentes de información, como la fotogenia, de forma que los algoritmos podrán predecir, a partir de una foto, la sexualidad de una persona (Wang&Kosinski, 2017), si es alcohólica (Yogeshwar, 2018), etc.; o incluso la voz, que permite averiguar si el dicente va a sufrir en cualquier momento un infarto (Comisión Europea, 2018b). El salto del gen a la imagen o al sonido nos muestra la verdadera dimensión del problema, ya que en realidad cualquier acto que realicemos podrá ser objeto de recopilación y análisis, fusionándose las pautas culturales y los elementos biológicos con objeto de *datificar* el comportamiento humano.

Consciente de este problema, el Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías, de la Comisión Europea, ha propuesto crear el derecho “a no ser perfilado, medido, analizado, aconsejado o *nudged*”⁴ (Comisión Europea, 2018a).

Sin embargo, la normativa europea parece que va en la dirección contraria. Así, los Reglamentos *e-Privacy* y de *libre circulación de datos no personales*, en tramitación en estos momentos, pretenden rentabilizar económicamente los datos generados por el internet de las cosas, las máquinas autónomas o los datos personales anonimizados (Escribano, 2018).

Además, la paradoja del derecho a no ser *datificado* es que los colectivos que no sean (*quieran ser, puedan ser*) objeto de minucioso escrutinio, serán objeto de discriminación negativa precisamente por su invisibilidad para los algoritmos, con consecuencias funestas incluso para su supervivencia. El derecho “a la criptografía” (Cotino, 2017) puede ser un suicidio aplicado a los colectivos.

En efecto, las predicciones de los Big Data basadas en muestras insuficientes, incompletas o no representativas de la sociedad podrían provocar la marginación de las comunidades cuya

⁴ “*Nudge* es el acto de intervenir o influenciar la toma de decisiones mediante sutiles reforzamientos positivos o negativos o sugerencias indirectas. *Nudge* en este contexto se podría traducir como estimular o inducir un comportamiento específico sin que el afectado sea consciente de eso.” Comisión Europea. Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías (2018a): “Statement on Artificial Intelligence, Robotics and ‘Autonomous’ Systems. Traducción de María José Villalobos.

información no haya podido ser recogida (v. gr. minorías étnicas o personas afectadas por enfermedades raras, Nuffield, 2018). O todos o ninguno, tal es la disyuntiva a que nos enfrentamos, ya que de otra forma se infringiría la tercera “v” de los Big Data, la que afecta a la variedad (las otras dos son *volumen* y *velocidad*).

Este problema ha sido puesto de manifiesto por la Agencia de la Unión Europea para los Derechos Fundamentales (FRA), al mostrar la contradicción que supone que en algunos países se prohíba la recolección de datos raciales o étnicos (no así en EEUU o Gran Bretaña) si, precisamente, para evitar la discriminación racial o étnica resulta imprescindible contar con estos datos (intimidad *versus* sesgo políticamente correcto *versus* discriminación positiva, tal es el trilema).

En resumen, genes, hábitos, rasgos faciales, redes sociales y familiares, etc., predecirán a la perfección nuestro comportamiento inmediato, aun cuando no seamos conscientes de nuestra personalidad o de nuestras decisiones. Esta pluralidad de fuentes de información personal creará perfiles que abarcarán tanto el rol de consumidores como el de votantes, pasando por el de trabajadores, amantes, progenitores o simples ciudadanos.

Obviamente, el manejo de un volumen de información *elefantiásico* como el descrito resulta singularmente perturbador, ya que confiere a quien lo detente un poder sin precedentes. Por ello, la Inteligencia Artificial, entendida como Big Data, debe ser controlada y regulada, con total independencia de su nivel real de “inteligencia” y despreciando problemas más metafísicos como su posible “autoconciencia”. Y la bioética no puede centrarse únicamente en la recopilación de los datos biológicos, minusvalorando el resto. Debemos asumir que el ser humano recién se ha convertido en un paquete de bits prestos a ser interpretados por un mercado globalizado, ya que el capitalismo financiero ha dado paso al capitalismo de los datos.

4. Inteligencia artificial como sesgo

Un análisis del significado exacto de término “algoritmo” excede nuestro trabajo. Si empleamos una definición sencilla, “algoritmo como conjunto de instrucciones”, tendríamos que distinguir entre algoritmos naturales (por ejemplo, las constantes cosmológicas, como la velocidad de la luz, serían las “instrucciones” que rigen nuestro universo, ya que no se pueden violar), y algoritmos creados por los seres humanos.

Este segundo grupo es el que nos interesa, ya que lo relevante de cualquier creación cultural humana es que proyectamos nuestra subjetividad, esto es, nuestros prejuicios, sobre la misma, con el inevitable resultado final de crear algoritmos con sesgos.

Pues bien, la cuestión clave en este caso es cómo programar con *sentido ético*, esto es, cómo conseguir que el conjunto de reglas que permiten ejecutar un programa no sean discriminatorias. Sin embargo, este objetivo es mucho más complejo de lo que pueda parecer a simple vista.

En efecto, a un programa se le exige que carezca de “sesgos”. Sin embargo, ¿cómo evitar los sesgos de los programadores⁵? Sostenía Gadamer que la Ilustración trajo consigo el prejuicio del “no prejuicio”, esto es, la creencia de que, en efecto, era posible no tener prejuicios.

Este problema explica que se trate de resolver la cuestión de los sesgos con conceptos ad hoc, como la “transparencia” que debe inspirar la programación de las IA (v. gr. gobierno francés, AA.VV., 2017a); la “comprensibilidad” (Executive Office of the President EEUU., 2016a:4), esto es, que los seres humanos afectados por una decisión basada en la Inteligencia Artificial puedan comprender, en un lenguaje inteligible, por qué el algoritmo toma esa decisión y no otra; o la “reversibilidad”, es decir, la posibilidad de poder cambiar dicha decisión (el algoritmo, en suma), con argumentos razonados (Biocat, 2017).

Sin embargo, la complejidad de la programación “ética” es enorme, de ahí que tratemos de clarificarla mostrando las cuatro alternativas posibles:

- a. La primera alternativa es que Inteligencia Artificial reproduzca milimétricamente los sesgos que ya existen en la sociedad, sin añadir ni corregir ninguno. Paradójicamente, esta es la concepción más humanizada de una IA, la que permite que actúe como lo haríamos nosotros. Cuando los investigadores del MIT elaboran una media probabilística de las decisiones que toman los seres humanos para resolver la “paradoja del tranvía” (Awad et al., 2018), pretenden implícitamente que los coches autónomos actúen como ser humano medio, prototípico o arquetípico. Si la media de la población atropellaría antes a un miembro de un determinado colectivo que a otro, el vehículo autónomo hará lo mismo.
- b. La segunda alternativa es que la IA actúe bajo el *velo de la ignorancia* de John Rawls. Para que la decisión sea justa, desprejuiciada y políticamente correcta, la IA no debe tener en cuenta factores étnicos, sociales, sexuales, de edad, etc. (“fair algorithms”, algo así como la *tábula rasa rousseauiana*).⁶ De esta forma, una IA neutral no es la que reproduce el

⁵ Un ejemplo nos lo proporciona el FRA europeo (2018), “Sin embargo, los humanos no describen las imágenes como neutrales. En concreto, un bebé con la piel blanca se describe como ‘bebé’, pero un bebé con la piel negra se describe como ‘bebé negro’”. (FRA, 2018).

⁶ El artículo 13.2 del Reglamento europeo de protección de datos, al regular la “información que deberá facilitarse cuando los datos personales se obtengan del interesado”, establece entre estas “la existencia de decisiones automatizadas, incluida la elaboración de perfiles (...) y, al menos en tales casos, información significativa sobre la lógica

comportamiento humano, sino la que se atropellaría incluso a sí misma porque estaría programada para no saber cuál será su rol en un hipotético accidente de tráfico (peatón, pasajero, conductor, etc.).

Sin embargo, los problemas que generan los paradigmas políticamente correctos, como el de Rawls (todo el mundo puede ser legislador, etc.), se pueden observar en el artículo 9.1 del novedoso Reglamento de Protección de Datos en la UE, que prohíbe que se empleen los Big Data para identificar “racial o étnicamente” a la población, aclarando, a continuación, que esta prohibición “no implica la aceptación por parte de la Unión de teorías que traten de determinar la existencia de razas humanas separadas” (apart. 51). ¿Comprenderá un algoritmo que las razas existen o no existen en función del contexto en que se emplee el término raza?

- c. La tercera alternativa es que la IA se erija en una especie de Mano Invisible de Adam Smith, esto es, en una secularización de la providencia divina agustina. El argumento es el siguiente: colectivamente debemos aceptar que la IA atropelle a alguno de nosotros porque estadísticamente los accidentes disminuyen (hay más accidentes si solo intervienen humanos que si participan las IA, Álvarez, 2017). De esta forma, el sacrificio de algunos miembros de la sociedad resulta beneficioso para el conjunto, aunque no seamos conscientes, ya que colectivamente se incrementa el índice de supervivencia (la información a la familia del fallecido y a su compañía de seguros habría que transmitirla precisamente bajo dicha cobertura justificativa: el deceso se produjo en aras del bienestar estadístico colectivo).

Esta tesis introduce a las IA en la idea del progreso ilustrado y del evolucionismo darwinista: se presupone que estamos ante una forma más civilizada de pauta cultural, donde los efectos colaterales se aceptan en aras de un bien mayor, el progreso indefinido.

- d. La cuarta alternativa es que la IA intervenga política y jurídicamente, esto es, que discrimine positivamente. Así, la Inteligencia Artificial debe corregir las desigualdades sociales favoreciendo a las minorías étnicas, sociales, etc. Sería una IA activista desde el punto de vista social; en su extremo, el *Hombre Nuevo* de Mao aplicado a la informática. Su misión es transformar la sociedad, haciendo tabla rasa con las pautas culturales discriminatorias. De esta forma, un motor de búsqueda que visibilice las contribuciones de las mujeres en ciencias sería una IA que cumpliría con el velo de la ignorancia de Rawls (sería neutral, si el 20% de las físicas en el siglo XX eran mujeres, el motor de búsqueda debe respetar esa

aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado”. Con razón el FRA se plantea si la información debe referirse a la racionalidad inherente a la técnica o a la técnica en sí misma.

proporción), esto es, en lenguaje político, una IA liberal; y un motor de búsqueda que priorice la aparición de mujeres para animar a las jóvenes a estudiar ciencias, donde están infrarrepresentadas, sería un motor que discrimina positivamente, esto es, un motor con conciencia de género. Algo similar se podría plantear con las minorías étnicas, religiosas (si introduces la palabra “Dios”, el motor de búsqueda colocaría en primer lugar a los jainistas, o a los zoroastristas, para evitar el monopolio judeocristiano o islámico de dicha palabra). Así, el algoritmo debe tener por objetivo cambiar la sociedad, con lo que el programador debe ser un militante político con conciencia social.

En conclusión, recién estamos comenzando a programar con gran incidencia en las relaciones sociales, étnicas, etc., sin que exista un consenso social acerca de cómo y para qué hacerlo. Entre el azar incontrolable, equiparable a la naturaleza, y el determinismo algorítmico, equiparable al determinismo genético, se sitúan las propuestas.

5. Inteligencia artificial como *cuestión sociolaboral*

El cuarto sentido que reseñamos atañe a un problema arraigado en los comienzos de la revolución industrial: cuanta más automatización, menos puestos de trabajo.

Las narrativas más agoreras predicen que los robots (autómatas, ordenadores, software, etc.) sustituirán a los humanos en casi todos los puestos de trabajo. El fantasma del ludismo (trabajadores emprendiéndolas a golpes con las máquinas) recorre la IA. El sector del transporte, perseguido por los vehículos autónomos, parece que será el primero afectado a gran escala, al que seguirán otros.

El problema sociolaboral nos conduce a sectores como la economía (sin trabajadores, no hay consumidores, ¿para quién producir, entonces?), de ahí que la sexta propuesta de la Declaración de Barcelona sobre IA, o el Parlamento europeo, sostenga que sustituir a los trabajadores por máquinas es un “error”; la política (la renta básica universal, que se discute en estos momentos, es una forma de anticiparse al problema de una generación de personas completamente desocupadas); o la fiscalidad y la seguridad social (Poquet, 2017), que plantean la duda de si los robots deben tributar y/o cotizar, bajo la idea, ciertamente inocente, de que un robot es un humanoide (v. gr. el Parlamento europeo hasta se ha molestado en describir qué es un robot, 2017). Esta ingenuidad es la que explica que nadie haya resuelto aún el problema de cómo imputar una declaración de la renta, o una nómina, a un algoritmo.

Las narrativas tecnoutópicas nos recuerdan cómo la desaparición de unos puestos de trabajo acarrió, en el pasado, la aparición de otros. Así, desde esta perspectiva, se crearán nuevos nichos de oportunidades que serán aprovechados por la parte de la clase trabajadora que sea capaz de reciclarse y adaptarse (*Quién se ha llevado mi queso*, de Spencer Johnson, sería el catecismo de este apostolado). Además, los trabajos monótonos, repetitivos y peligrosos serán desempeñados por las máquinas, mientras que los trabajos creativos, imaginativos, etc., serán reservados a los humanos.

Esta narrativa no resuelve el problema de qué hacer con la parte de la población *no reprogramable*, ni cómo redistribuir la riqueza en una sociedad hipertecnificada, pero, sobre todo, si realmente podemos establecer paralelismos entre el tránsito de la sociedad agrícola a la industrial, y el de la industrial a una sociedad regida por las tecnologías disruptivas (biotecnología, nanotecnología, Inteligencia Artificial, biología sintética, etc.).

Sin embargo, y a pesar de la gravedad de los problemas expuestos, la cuestión sociolaboral constituye una fractal de temáticas mucho más complejas, como nuestra relación psicoafectiva con los robots. En efecto, la proyección de sentimientos o afectos sobre estos (nadie empatiza con una lavadora si es destrozada, pero la reacción difiere si estamos ante un humanoide), a medio camino entre la antropomorfización y la aversión (el famoso “valle inquietante”), con el telón de fondo de la cuestión sexual (que significativamente ha sido obviada de los dictámenes e informes que estamos citando), nos permite augurar que en los próximos años los puestos de trabajo disminuirán de forma inversamente proporcional a como crecerán las terapias psicoanalíticas.

Estas cuestiones justifican expresiones como la del “contacto humano significativo”, del Grupo Europeo de Ética de las Ciencias (CE. 2018a). Dicho grupo ha propuesto que la interacción sea voluntaria, consciente (la persona debe saber que está ante una IA) y limitada (no todos los procesos se pueden delegar en una máquina), con objeto de dignificar la relación humano/máquina. Significativamente, la propuesta remarca la relevancia de la dignidad de forma análoga a como el convenio europeo de bioética separaba, con objeto de destacarla, la dignidad de los derechos humanos.⁷

En resumen, la cuestión de fondo es si ya hemos pasado por esta situación (v. gr., en el siglo XVIII) o, por el contrario, estamos ante los umbrales de una nueva era donde los humanos, salvo los

⁷ Convenio de 4 abril 1997, ratificado por Instrumento de 23 julio 1999. RCL 1999\2638. Instrumento de Ratificación del Convenio para la protección de los derechos humanos y la dignidad del ser humano con respecto a las aplicaciones de la Biología y la Medicina, Oviedo, 4 de abril de 1997.

propietarios de las IA y el resto de tecnologías exponenciales, simplemente sobramos, no ya bajo la figura del trabajador/consumidor que sostiene el capitalismo actual, sino como ciudadanos.

6. Inteligencia artificial como ente sin conciencia

El temor a la Inteligencia Artificial se justifica, no porque este tipo de entidades piensen, razonen o puedan equipararse de alguna manera a nuestras habilidades cognitivas, sino porque no es necesario que sean conscientes para que los riesgos sean extremos.

En efecto, a la hora de prevenir los peligros inherentes a la IA siempre rememoramos los humanoides de las novelas y películas de ciencia ficción. Sin embargo, como los avances en conciencia artificial son nulos en estos momentos, el resultado es que consideramos estos escenarios como algo ficticio, lo que nos lleva a minimizar los riesgos. Si un ordenador no puede razonar, ¿por qué temerle?; si la conciencia solo es una propiedad humana, por mucho que los etólogos señalen a otras especies vivas y los antropólogos cuestionen nuestra suficiencia antropocéntrica, ¿por qué va a suponer un riesgo una entidad electrónica que ni siquiera sabe que existe?

Pues bien, aquí es donde entra en juego el concepto de “automaticidad”, clave en el presente epígrafe. No hace falta que una IA piense, razonen ni tome conciencia de sí misma para que sea peligrosa para la especie humana. Basta con que *actúe*.

El Grupo Europeo sobre Ética de la Ciencia, de la Comisión Europea, reflexionaba en un reciente informe sobre un evento que ha pasado desapercibido (probablemente porque es menos mediático que el ajedrez en occidente): “It is impossible to understand how exactly AlphaGo managed to beat the human Go World champion (...) In this sense, their actions are often no longer intelligible, and no longer open to scrutiny by humans” (CE. 2018a). Como podemos observar, la perplejidad de dicha institución no radica ni en la victoria ni en la rapidez con que el programa se instruyó, sino en la forma en que tomó decisiones *autónomas* para ganar de una forma tan aplastante.

Otro ejemplo en este sentido nos lo ofrece el incidente de dos ordenadores de Facebook conectados entre sí para desarrollar una determinada tarea. Desarrollaron un lenguaje propio que, a ojos de los programadores, resultaba incomprensible (la rápida desconexión de los ordenadores no sabemos si obedeció a la ininteligibilidad o a la imprevisibilidad). Sin embargo, lo que realmente había sucedido es que optimizaron el lenguaje en que fueron programados para

acelerar la tarea encomendada (Ercilla, 2018a). Es decir, los ordenadores tomaron una ruta no programada, incomprensible e imprevisible para nuestra especie.

Con todo, el mejor ejemplo de “automaticidad inconsciente” lo constituye la propia naturaleza. En efecto, tal y como señalara con perspicacia Gould,⁸ a la naturaleza no le ha hecho falta voluntariedad ni conciencia para crear *algo* como nosotros. Así, el *algoritmo* con el que ha trabajado la naturaleza está compuesto de hidrógeno, tiempo, gravedad, constantes cosmológicas, etc., con un resultado final ciertamente sorprendente, al menos desde la perspectiva humana, hasta el punto de que no sabemos si es una consecuencia necesaria del universo en el que vivimos (principio antropocéntrico fuerte) o algo completamente azaroso, singular, e incluso único (v. gr., no hay más vida inteligente en el Universo que nosotros).

Pues bien, una Inteligencia Artificial que operara a ciegas, esto es, sin ningún objetivo prediseñado, podría, si se diseñara con el nivel de complejidad adecuado, provocar cambios estructurales en la civilización humana sin que pudiéramos, una vez iniciado el proceso, detenerla. Es como si intuyéramos que los algoritmos inicialmente dispuestos en una IA pueden, en cierta manera, *cobrar vida* y tomar direcciones inimaginables a priori (los casos de AlphaGo y de Facebook serían dos incipientes ejemplos de esta ausencia de control final).

La expresión “caja negra” (“una máquina opaca en la que se introducen inputs y se generan, mediante un proceso en cierta forma indescifrable, outputs sin explicación posible”, Executive, 2016b) alude precisamente a este problema).

El temor a conceder una plena autonomía a una IA explica que la Declaración de Barcelona (Biocat, 2017), sobre Inteligencia Artificial, haya propuesto una “autonomía limitada” para evitar escenarios fuera de control (quinta propuesta), de forma que todo sistema de IA que entrañe riesgos críticos dependa, en última instancia, de *algún* humano;⁹ que el informe de Nuffield destaque los riesgos de los ataques a gran escala que se pueden producir precisamente por procesos automatizados (2018); que en España se prohíba la circulación de vehículos

⁸ “Si la creación exige un creador visionario, entonces ¿cómo logra la evolución ciega construir cosas nuevas tan espléndidas como nosotros mismos?” GOULD, S. J. (1996): “Creating the Creators”, en *Discover Magazine*, Octubre, pp. 42-54, en LÓPEZ DE MANTARÁS, 2017.

⁹ Armengol et al. (2018) destacan cómo la colusión tácita entre algoritmos puede afectar a la libre competencia.

completamente autónomos, salvo en fase de pruebas,¹⁰ y que se requiera una definición clara y exacta de qué son los “sistemas de decisión automatizada” (Resiman et al., 2018).¹¹

Por último, la autonomía de las IA afecta a dos conceptos claves, “responsabilidad” e “irreversibilidad”, de ahí que el Grupo Europeo de Ciencias haya propuesto distinguir entre “autonomía” y “automatismo” (CE 2018a), con objeto de mostrar que las IA pueden estar bajo la cobertura de la “automaticidad”, pero que esto no implica que sean entidades propiamente autónomas, ya que la autonomía presupone “responsabilidad moral”, rasgo ausente en este tipo de entidades.¹²

En conclusión, los algoritmos que rigen la IA pueden regirse por reglas que desencadenen decisiones (“desencadenar decisiones” no es lo mismo que “tomar decisiones”, ya que esta última acción implica conciencia, mientras que aquella implica automatización) con resultados incontrolables, impredecibles y estructurales, de ahí los esfuerzos normativos para asegurar que el control último sea de naturaleza humana.

7. Inteligencia artificial como ente con conciencia

En estos momentos se discute si los robots pueden ser equiparados a las personas (en realidad los igualaríamos a los animales, ya que presuponemos que su nivel de inteligencia será, en todo caso, inferior al nuestro¹³); si pueden gozar al menos de personalidad jurídica, constructo análogo al de las sociedades mercantiles, de gran utilidad en un contexto capitalista (Parlamento europeo, 2017; Ercilla, 2018a; González, 2016);¹⁴ si podemos establecer relaciones afectivas con este tipo de entidades (algo así como una mascota electrónica sofisticada, por encima de un hámster, pero por debajo de un perro); si pueden satisfacer necesidades sexuales (esta temática es todo un mundo en sí misma), etc. Pero ninguna de estas discusiones, que pueden llegar a ser realmente

¹⁰ Dirección General de Tráfico española (Instrucción 15/V-113, de 12 de noviembre de 2018), en Álvarez, 2018

¹¹ Citando como precedente el artículo 4 del Reglamento europeo de protección de datos, en especial los apartados uno y dos.

¹² Carmen Nuñez (2018) muestra en su artículo cómo el sistema de responsabilidad civil atribuible a los robots se ha quedado desfasado.

¹³ En robótica, la traducción jurídica de la “naturaleza humana” implicaría recopilar las características que hacen que un individuo sea sujeto de derechos y no un mero objeto de los mismos. (Santamaría, 2006):

¹⁴ Un punto de vista intermedio entre estas dos cuestiones es plantearse si pueden ser imputables desde el punto de vista penal (De la Cuesta, 2016).

fascinantes, se enfrenta al verdadero problema de fondo: ¿puede una Inteligencia Artificial cobrar conciencia *de sí misma*, superándonos en capacidades cognitivas?¹⁵

A priori, parece razonable pensar que nunca se alcanzará el nivel de complejidad humana (millones de años de evolución azarosa no pueden ser replicados, menos aún superados, informáticamente). Sin embargo, no lo sabemos con la suficiente certeza, y de hecho, nadie se atreve a pronosticar qué puede suceder en las próximas décadas. El problema latente es que, además, no nos podemos equivocar en esta cuestión, ya que podría suponer un riesgo existencial para nuestra especie.

Estas cautelas se han puesto de manifiesto en el informe del organismo que asesora a la presidencia norteamericana: “*es poco probable* que la inteligencia artificial iguale o supere a los seres humanos en los próximos *veinte años*” (Executive Office of the President 2016a); así como en el Parlamento europeo: “Considerando que existe la posibilidad de que a largo plazo la inteligencia artificial llegue a superar la capacidad intelectual humana” (2017).

Para abordar estos hipotéticos escenarios se están creando una serie de expresiones que tratan de no alarmar excesivamente a la opinión pública (en este caso, para evitar regulación y/o control, esto es, la estrategia inversa a la que analizamos en la IA como *trending topic*), hasta el punto de que los investigadores han desarrollado una neolengua propia, incomprensible para el profano. Así, sus expresiones cumplen la doble función de ser eufemísticas (el equivalente al lenguaje políticamente correcto en las cuestiones sociales controvertidas), y servir, a la vez, de advertencia, aunque sea entrelíneas, de que estamos ante algo *diferente*. En este sentido, podemos destacar sintagmas como “Inteligencia Artificial *Fuerte*” (el resto se supone que es *débil*); “Aprendizaje *Profundo*” (la IA no solo procesa información algorítmica, sino que además aprende, retroalimentándose con su entorno, de donde podemos inferir que la mera programación sería aprendizaje *superficial*); “mecatrónica” (una mezcla de IA y aprendizaje profundo, lo que presupone que son cosas diferentes, CE, 2018), o se contraponen dicotomías como *Data Driven AI* (cualquier IA con menos de diez años) y *Knowledge-based AI* (que serían los hardware de la década de los setenta). Con tales juegos lingüísticos se pretende diferenciar cualquier tarea que pueda ser implementada por un programa informático, de las capacidades inherentes a una entidad con conciencia de sí misma.

Pues bien, el nivel de complejidad de la cuestión que estamos analizando nos lleva a sostener que en realidad estamos ante tres preguntas diferentes, aunque interrelacionadas: a) ¿es

¹⁵ Una Orden del Ministerio de Economía, de 23 de enero de 1990, definía la Inteligencia Artificial en relación a las “funciones asociadas normalmente con la percepción y el razonamiento o aprendizaje humanos”.

posible emular artificialmente el cerebro humano, es decir, replicarlo en otro soporte, como el silicio, a base de algoritmos?; b) ¿es posible crear un ente inteligente, con capacidad para interactuar con su entorno, aprender, y plantearse objetivos y metas, *aunque esta forma de inteligencia no tenga nada que ver con la nuestra*. En este caso sería algo así como crear vida, solo que, además de superar el reto de crearla, esta debe ser inteligente; y mientras los biólogos se centran en desarrollar vida siguiendo los supuestos pasos que hemos seguidos nosotros (sopa primordial, evolución, tiempo, etc.), los informáticos estarían haciendo lo mismo pero centrándose únicamente en los algoritmos que procesan la inteligencia. Cabe plantearse, además, si estas dos formas de búsqueda de la vida pueden confluír o no; c) ¿es posible crear un ente electrónico *con conciencia de sí mismo*? En realidad, no hace falta que dicho ente sea realmente inteligente, bastaría con que tuviera conciencia para que el logro realmente nos cautivara, sorprendiera y, lógicamente, alarmara.

Y estas tres cuestiones interrelacionadas se pueden reconducir a un problema de naturaleza cosmológica: ¿qué es *pensar*? En efecto, presuponemos que la única manera de “pensar” es la propia de nuestra especie, o, al menos, de determinados seres vivos basados en la misma química que nosotros, pero, ¿se puede *pensar* empleando otros elementos químicos (como el silicio), otros soportes (como los chips); y otro tipo de sistemas de codificación y transmisión de la información (como el sistema digital de los ordenadores)? La realidad es que no lo sabemos.

El informe norteamericano reconoce la imposibilidad de definir qué es una IA, de ahí que se haya centrado en las habilidades que debe tener para que se considere, en efecto, *inteligente*. Ante esta dificultad, el Parlamento europeo ha optado por definir, no una IA, sino un robot (2017), como si el problema fuesen los humanoides y no los algoritmos. Pues bien, lo que nos demuestran estos intentos de acotar de qué estamos hablando es que el patrón de medida es completamente antropocéntrico: una IA piensa o es consciente si es como nosotros; sin embargo, ¿acaso no sería una prodigiosa hazaña que una IA pensara, al menos, como un perro, un simio o un erectus?

La vida humana se basa en el carbono. Nuestro cerebro es el producto del azar y de la selección natural (paradigma darwinista). Hasta ahora, no hemos sido capaces de crear vida desde la nada (el famoso experimento de Muller solo logró aminoácidos), ni siquiera disponiendo de la capacidad de procesamiento de los ordenadores, de la energía que podemos generar, y de la información que tenemos acerca del funcionamiento de los genes (Barrera&López Baroni, 2018). Tampoco hemos localizado ningún signo de vida extraterrestre, a pesar de la trivial composición de nuestros organismos (carbono, oxígeno, nitrógeno, etc.), que en principio nos llevaría a pensar que la vida debería haberse producido una y otra vez, al menos en nuestra galaxia.

La pregunta clave es la siguiente: una Inteligencia Artificial ¿es una forma de vida *en sí misma*? Si es posible crearla, otras formas de vida extraterrestre inteligente ya lo habrían hecho (han dispuesto de millones de años para ello) ¿Por qué no recibimos ningún indicio de la existencia de esta forma de vida que razonablemente debería haberse producido una y otra vez (toda forma de vida orgánica, pasado un determinado estadio de civilización, debería desarrollarla)? ¿Significa eso que es imposible de crear, y por eso, no debemos preocuparnos? ¿O implica, por el contrario, que la ausencia de señales de vida extraterrestre de composición orgánica se debe precisamente a que cruzaron el umbral de la inteligencia artificial (el Gran Hiato de Bostrom, 2016, podría ser justo ese)? Estas reflexiones explican que López de Mantarás (2017) sitúe el nivel de dificultad de estas temáticas en el mismo plano que otros interrogantes científicos, como el origen del universo, de la vida, o de la propia estructura del universo.

Pues bien, el carácter etéreo, metafísico, de estas cuestiones es lo que justifica que el derecho no preste atención a estos hipotéticos escenarios.¹⁶ No es necesario regular lo imposible ni lo altamente improbable (para qué vamos a prohibir penalmente, por ejemplo, enviar señales al espacio, si la probabilidad de que tenga alguna consecuencia es irrelevante).¹⁷ Por ese motivo, las reflexiones acerca de una Inteligencia Artificial de esta naturaleza pertenecen únicamente al plano académico, y rara vez trascienden al mundo político/legislativo, lo que no impide que en los informes o dictámenes cada vez aparezcan más reflexiones sobre esta posibilidad.

8. Inteligencia artificial como disciplina convergente

Desde esta perspectiva, lo relevante no es la Inteligencia Artificial en sí misma, sino la forma en que dicha disciplina interacciona y se retroalimenta con otras, con un resultado final difícil de predecir.¹⁸

Estas tecnologías son la biotecnología, la biología sintética, la edición genómica, la nanotecnología y la cognotecnología, que, junto a la Inteligencia Artificial y los Big Data, se están agrupando con acrónimos más o menos afortunados que pretenden mostrar cómo debemos

¹⁶ Ercilla denomina “neoderecho” a esta hipotética regulación (2018b).

¹⁷ El artículo de Laurent Orseau y Stuart Armstrong, *Safely Interruptible Agents*, explora precisamente esta cuestión, el “botón rojo” para una IA (Google DeepMind y The Future of Humanity Institute, junio de 2016, Machine Intelligence Research Institute).

¹⁸ La primera vez que se citó la Inteligencia Artificial en la Unión Europea fue en relación al “reconocimiento de formas” (Decisión del Consejo de 26 de noviembre de 1985, Acción COST 13. DOUE-L-1985-80993)

adoptar una perspectiva holística, donde el todo es más que la suma de las partes, con objeto de comprender los desafíos contemporáneos.

Así, las expresiones suelen variar entre lo neutral (Tecnologías convergentes); lo políticamente correcto (Tecnologías Disruptivas, denominación que hace hincapié en la capacidad para modificar estructuralmente a la civilización humana); las analogías cosmológicas (BANG, acrónimo de Bits, Átomos, Neuronas y Genes, en relación a la expresión Big Bang); lo utópico (Tecnologías Exponenciales, por su potencial capacidad para aumentar exponencialmente las capacidades humanas; tecnochovinismo, en sentido crítico, Broussard, 2018); o lo filosófico (Tecnologías de la Sospecha, en referencia a Ricoeur y a los filósofos que cuestionaron los fundamentos de la sociedad occidental, López Baroni, 2018).

Quizá el documento fundacional de esta forma de analizar la complejidad sea un informe norteamericano titulado *Converging Technologies for Improving Human Performance* (Mihail C. Roco&Sims Bainbridge, William, *National Science Foundation*), de 2003, al que siguió otro europeo, de 2004, *Converging Technologies – Shaping the Future of European Societies* (Nordmann et al, Comisión Europea). A diferencia del norteamericano, en la portada del europeo figura la expresión “Nano-Bio-Info-Congo-Socio-Anthro-Philo-Geo-Eco-Urbo-Orbo-Macro-Micro-Nano-Geo-Eco-Micro-Nano”, como si fuese un catálogo de las esferas humanas que se verán afectadas por este proceso de convergencia.^{19,20}

El núcleo de este tipo de reflexiones es que si aprendemos a controlar la materia a escala atómica y/o molecular, a modificar, alterar o transferir genes mediante las modernas técnicas de edición genómica, comprendemos el funcionamiento del cerebro, y por tanto, de la conciencia, y diseñamos adecuadamente los algoritmos, tanto para apoyar aquellas tecnologías como para capacitar la Inteligencia Artificial de forma mimética a nuestra especie, el resultado puede suponer un hiato en la existencia de la vida tal y como la conocemos en el planeta.

¹⁹ Para valorar estas convergencias podemos recordar la coincidencia, en poco espacio de tiempo, de una serie de avances científicos repercusión global, En efecto, el primer mensaje que se envió a través de internet (Lawrence Roberts, Vinton Cerf y Tim Berners-Lee, Premio Príncipe de Asturias, 2002, por sus aportaciones) coincide con el año en que llegamos a la Luna, y precede en tan solo cuatro años al nacimiento de la biotecnología moderna (1973, ADN recombinante, Cohen, Berg y otros); la expresión Inteligencia Artificial no se acuñó hasta 1956, a partir del artículo de Alang Turing “Computing Machinery and Intelligency” (Executive 2016a:5); y en estos momentos asistimos, estupefactos, a una verdadera explosión de la edición genómica con la aparición de la técnica CRISPR.

²⁰ En este sentido, el Real Decreto 1008/2017, de 1 de diciembre, destaca la relevancia que tiene para la seguridad nacional la interacción del internet de las cosas, la inteligencia artificial, la ingeniería genética y la robotización.

Desde el punto de vista del derecho, debemos destacar cómo el sistema de patentes sobre las entidades vivas producidas por la biotecnología (López Baroni, 2017), plantea desafíos parecidos a las patentes sobre los algoritmos, los programas de ordenador (Bustillo, 2008), y, en definitiva, sobre la Inteligencia Artificial.

En efecto, originalmente, las patentes no estaban previstas ni para la materia viva, ni para los algoritmos, que se excluían del proceso de apropiación humano por ser productos de la naturaleza (descubiertos y, por tanto, *no inventados*). Sin embargo, los avances en Biotecnología y en Inteligencia Artificial han forzado que los productos de estas disciplinas se hayan incorporado a los sistemas de protección de la propiedad industrial.

En resumen, se intuye que las tecnologías disruptivas nos van a obligar a reescribir la ley de Amara en un sentido mucho más desasosegante:²¹ no solo subestimamos los efectos de la tecnología a corto plazo, sino que ni siquiera somos capaces de pronosticar, predecir o intuir sus efectos a largo plazo. Sin duda, el estudio integrado de las tecnociencias se ve dificultado tanto por la celeridad con que avanzan, como por la nebulosa que envuelve la forma en que interaccionan y se retroalimentan estas disciplinas. Pero estudiarlas por separado puede suponer un acto de temeridad colectiva.

9. Conclusiones

Al igual que sucede con el resto de tecnologías disruptivas, sobre la Inteligencia Artificial proyectamos tanto nuestros miedos (una IA consciente y todopoderosa que nos extermine), como nuestras esperanzas (un soporte en el que volcar nuestra identidad y evitar la muerte).

Sin embargo, todo apunta a que la IA, a pesar de la cantidad de adjetivos que la rodean (profunda, etc.), se reducirá en las próximas décadas a mejoras en el procesamiento de la información. Por inimaginables e impredecibles que sean los resultados de estos avances, difícilmente nos autorizarán a considerar a dichas entidades como organismos vivos, menos aún con conciencia de sí mismos.

Sin embargo, el hecho de que no tengamos una absoluta seguridad sobre las limitaciones inherentes a tratar de replicar y mejorar lo que se ha generado por azar durante millones de años, esto es, la vida, debe llevarnos a extremar las precauciones, y por ende, el rigor conceptual.

²¹ La ley dice que sobrestimamos los efectos de la tecnología a corto plazo, pero los subestimamos a largo plazo.

De hecho, la confusión que rodea a la expresión Inteligencia Artificial desdibuja, cuando no frivoliza, los riesgos. La incipiente normativa, fraccionada y dispersa, tan pronto se hace eco de las hipótesis más tremendistas como trata de regular la recolección de datos triviales, prueba de la desorientación que embarga al legislador ante la avalancha de hipotéticos escenarios.

Por último, lo más prudente es tratar la inteligencia artificial como una fractal de una realidad mucho más compleja, la compuesta por la interacción de las tecnologías disruptivas. Sin embargo, la paradoja de este ejercicio de sensatez es que aumenta exponencialmente la dificultad a la hora de su regulación y control. Parafraseando a Kant, podríamos decir que poco a poco nos estamos internando en el oscuro océano, sin costa ni faros, de la tecnometafísica.

Bibliografía

- ◆ AAVV. (2017a): *Rapport de synthese. France Intelligence Articielle*. Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche.
- ◆ AAVV. (2017b): *Asilomar AI principles*. Asilomar Conference. Future of Life Institute.
- ◆ ÁLVAREZ OLALLA, Pilar (2017): "Desafíos legales ante la circulación de los coches autónomos: implicaciones éticas, responsabilidad por accidente y ciberseguridad". *Revista Doctrinal Aranzadi Civil-Mercantil*, núm. 2/2017. BIB 2017/10732.
- ◆ ARMENGOL I GUSULL, Oriol; DE BLAS AGUILERA, Jaime; VIGÓN UZQUIANO, Inmaculada (2018): "Algoritmos y Derecho de la Competencia". *Anuario de Derecho de la Competencia*. Aranzadi. Bib 2018/10276.
- ◆ AWAD, Edmong; DSOUZA, Sohan; KIM, Richard, SCHULZ, Jonathan; HENRICH, Joseph; SHARIFF, Azim; BONNEFON, Jean-François; RAHWAN, Lyad. (2018): "The Moral Machine experiment". *Nature*, 563, pp. 59-64.
- ◆ BARRERA GARCÍA, Alberto & LÓPEZ BARONI, Manuel Jesús, (2018): ¿Existe una frontera nítida entre la materia viva y la materia inerte? *Revista de Bioética y Derecho*, núm. 43, pp. 9-31.
- ◆ BIOCAT Y OBRA SOCIAL "LA CAIXA" (2017): *Declaración de Barcelona para un desarrollo y uso adecuados de la inteligencia artificial en Europa*. 8 de marzo.
- ◆ BUSTILLO SÁIZ, María del Mar (2008): "Hacia la patentabilidad de los programas de ordenador: un diálogo particular entre el Derecho y la Economía (II)". *Revista Aranzadi de Derecho y Nuevas Tecnologías*, núm. 16/2008.

- ◆ BOSTROM, Nick, (2016): *Superinteligencia: Caminos, peligros, estrategias*. Teell.
- ◆ BROUSSARD Meredith (2018): *Artificial Unintelligence. How Computers Misunderstand the World*. The MIT Press.
- ◆ COMISIÓN EUROPEA. GRUPO EUROPEO SOBRE ÉTICA DE LA CIENCIA Y LAS NUEVAS TECNOLOGÍAS. (2018a): “Statement on Artificial Intelligence, Robotics and ‘Autonomous’ Systems, European Group on Ethics in Science and New Technologies”.
- ◆ COMISIÓN EUROPEA, COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES (2018b) Inteligencia artificial para Europa {SWD(2018) 137 final}. Bruselas, 25.4.2018 COM, 237 final.
- ◆ COMITÉ ECONÓMICO Y SOCIAL EUROPEO (2017): Dictamen sobre la inteligencia artificial: las consecuencias de la inteligencia artificial para el mercado único (digital), la producción, el consumo, el empleo y la sociedad. Dictamen de iniciativa. Pleno del CESE de 31 de mayo de 1 de junio de 2017. Diario Oficial de la Unión Europea (31-08-2017).
- ◆ COTINO HUESO, Lorenzo (2017): “Big data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales”. *Dilemata*. Nº 24, pp. 131-150.
- ◆ COUNCIL OF EUROPE (2017a), *Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data*, T-PD(2017) 01, 23 January.
- ◆ COUNCIL OF EUROPE (2017b): *Study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications*. Committee of experts on internet intermediaries, MSI-NET (2016) 06 rev6.
- ◆ DE LA CUESTA AGUADO, Paz (2016): “La ambigüedad no es programable, racionalización normativa y control interno en inteligencia artificial”. *Revista Aranzadi de Derecho y Proceso Penal* num.44/2016. BIB 2016/85668.
- ◆ DIAMANDIS, Peter; KOTLER, Steven (2013): *Abundancia: el futuro es mejor de lo que piensas*. Antoni Bosch Ed.
- ◆ ERCILLA GARCÍA, Javier:
 - (2018a): “Aproximación a una Personalidad Jurídica Específica para los robots”, *Revista Aranzadi de Derecho y Nuevas Tecnologías*, núm. 47/2018, BIB 2018/11197.
 - (2018b): *Normas de derecho civil y robótica. Robots inteligentes, personalidad jurídica, responsabilidad civil y regulación*. Aranzadi.

- ◆ ESCRIBANO CAÑAS, Blanca (2018): “Dos proyectos de Reglamento europeos mantienen en vilo a la llamada economía de los datos de la Unión”. *Actualidad Jurídica Aranzadi*. Núm. 937/2018.
- ◆ EUROPEAN PARLIAMENT:
 - (2017a), *Fundamental rights implications of big data*, P8_TA-PROV (2017)0076.
 - (2017b), *Civil Law Rules on Robotics*, P8_TA(2017)0051.
- ◆ EXECUTIVE OFFICE OF THE PRESIDENT NATIONAL SCIENCE AND TECHNOLOGY COUNCIL COMMITTEE ON TECHNOLOGY:
 - (2016a): *Preparing for the future of artificial intelligence*.
 - (2016b): *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*.
- ◆ GONZÁLEZ GRANADOS, Javier, (2016): “Derecho y robots en la Unión Europea: hacia una persona electrónica. En <https://tallerdederechos.com/derecho-y-robots-en-la-union-europea-hacia-una-persona-electronica/>
- ◆ FRA. European Union Agency for Fundamental Rights (2018): *Big Data: Discrimination in data-supported decision making*.
- ◆ KURZWEIL, Kurz (2012): *La singularidad está cerca: cuando los humanos trascendamos la biología*. Lola Books.
- ◆ YOGESHWAR, Ranga (2018): Entrevista. Irene Hernández Velasco. El Mundo. 22/10.
- ◆ LÓPEZ DE MANTARÁS, Ramón (2017): “Ética en la inteligencia artificial”, *Investigación y Ciencia*. Agosto. P. 40.
- ◆ LÓPEZ BARONI, Manuel Jesús (2018) “El criterio de demarcación en las Biopatentes.” *Anales de la Cátedra Francisco Suárez*, 52, pp. 131-153.
- ◆ LÓPEZ BARONI, Manuel Jesús (2018): “Las narrativas en biotecnología”. En *Argumentos de Razón Técnica*, 21, pp. 47-76.
- ◆ NUFFIELD COUNCIL ON BIOETHICS (2018): *Artificial intelligence (AI) in healthcare and research*.
- ◆ NUÑEZ ZORRILLA, María Carmen (2018): “Los nuevos retos de la Unión Europea en la regulación de la responsabilidad civil por los daños causados por la inteligencia artificial”. *Revista Española de Derecho Europeo*. BIB 2018/9706.

- ◆ NORDMANN, Alfred et al. (2004): *Converging Technologies – Shaping the Future of European Societies*.
- ◆ O’NEIL, Cathy (2018): *Armas de destrucción matemática. Cómo el big data aumenta la desigualdad y amenaza la democracia*. Traducción de Violeta Arranz de la Torre. Capitán Swing.
- ◆ ORSEAN, Laurent; ARMSTRONG, Stuart (2016): “Safety Interruptible Agents” Machine Intelligence Research Institute.
- ◆ PARLAMENTO EUROPEO (2017). Resolución 2015/2103.INL del Parlamento Europeo, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica.
- ◆ POQUET CATALÁ, Raquel (2017): “Interrogantes abiertos de la robótica en el Derecho del Trabajo”. *Revista de Información Laboral*, núm. 11/2017. BIB2017/43180.
- ◆ RESIMAN, Dillon; SCHULTZ, Jason; CRAWFORD, Kate; WHITTAKER, Meredith (2018): “Algorithmic Impact Assessments: A practical framework for public agency accountability”. AI Now Institute.
- ◆ RESOLUCIÓN DEL PARLAMENTO EUROPEO con recomendaciones destinadas a la Comisión sobre normas de Derecho Civil sobre robótica.
- ◆ SANTAMARÍA IBEAS, Javier (2006): “Blade Runner y la titularidad de los derechos fundamentales”. *Revista Aranzadi de Derecho de Deporte y Entretenimiento*. BIB 2006/159.
- ◆ WANG, Yilun; KOSINSKI, Michal (2018): “Deep neural networks are more accurate than human at detecting sexual orientation from facial images”. *Journal of Personality and Social Psychology*, Vol. 114, Issue 2, pp. 246-257.

Fecha de recepción: 1 de diciembre de 2018

Fecha de aceptación: 31 de enero de 2019