

De la persona a la personalidad algorítmica

A propósito de la personalidad
jurídica de la inteligencia artificial

Javier González Granado



Organització
de les Nacions Unides
per a l'Educació,
la Ciència i la Cultura



Càtedra UNESCO de Bioètica
de la Universitat de Barcelona



Observatori de
Bioètica i Dret
Universitat de Barcelona

De la persona a la personalidad algorítmica

De la persona a la personalidad algorítmica

A propósito de la personalidad
jurídica de la inteligencia artificial

Javier González Granado



UNIVERSITAT DE
BARCELONA

Edicions



Observatori de
Bioètica i Dret
Universitat de Barcelona

© Edicions de la Universitat de Barcelona

Adolf Florensa, s/n

08028 Barcelona

Tel.: 934 035 430

Fax: 934 035 531

comercial.edicions@ub.edu

www.publicacions.ub.edu



ISBN

978-84-9168-602-6

Esta obra se inscribe en el proyecto de investigación DER2017-85174-P, financiado por el Ministerio de Economía y Competitividad, Programa Nacional de ciencias sociales, económicas y jurídicas.



Este documento está sujeto a la licencia de Reconocimiento-NoComercial-SinObraDerivada de Creative Commons, cuyo texto está disponible en: <http://creativecommons.org/licenses/by-nc-nd/4.0/>.



Sumario

Presentación, por Antonio Garrigues Walker.....	11
Prólogo, por Itziar de Lecuona	13
Introducción	25
 1. ORÍGENES.....	29
1.1. <i>Homo sapiens, Homo juridicus</i>	29
1.1.1. Hordas, tribus, jefaturas y Estados.....	29
1.1.2. Sociedades prejurídicas	30
1.1.3. Sociedades protojurídicas	32
1.1.4. Sociedades jurídicas plenas	32
1.2. Los destinatarios de las primeras normas jurídicas.....	33
1.2.1. Asia Menor, Egipto y Grecia	33
1.2.2. Roma	35
1.3. La concreción del concepto	36
 2. SER HUMANO Y SUJETO DE DERECHOS: RAZONES, DEFECTOS Y EXCESOS	41
2.1. Razones	42
2.1.1. La comunidad moral	42
2.1.2. Contractualismo social.....	42
2.1.3. Utilitarismo	45
2.2. Defectos.....	47
2.2.1. Esclavos, mujeres y enemigos	47
2.2.2. Embriones.....	49
2.2.3. Animales	49
2.3. Excesos.....	50
2.3.1. Personas jurídicas	50
2.3.2. Personas electrónicas	51
 3. HACIA UNA PERSONALIDAD ALGORÍTMICA	53
3.1. Robots e inteligencia artificial.....	53
3.1.1. Del mito a la realidad	53

3.1.2. De qué hablamos cuando hablamos de robots	56
3.2. Roboética y derecho de los robots	58
3.3. La situación legal actual	59
3.3.1. <i>Lege data</i>	59
3.3.2. <i>Lege ferenda</i>	61
3.4. ¿Por qué personas? Pros y contras.....	65
3.4.1. El agente moral.....	65
3.4.2. Agentes morales implícitos, explícitos y plenos.....	66
3.4.3. Características del agente moral: autonomía, interactividad y adaptabilidad.....	67
3.4.4. Una moralidad sin mente	70
3.4.5. Moralidad distribuida: la complejidad como causa.....	71
3.4.6. La cuestión de la responsabilidad jurídica	73
3.4.7. Razones para una negativa	76
3.4.7.1. Las objeciones a su participación en la toma de decisiones con alcance moral	76
3.4.7.2. La carta abierta de la sociedad civil a los gobernantes de la Unión Europea	76
3.4.7.3. El <i>Estudio del Departamento de Política de Derechos de los Ciudadanos, Libertades Fundamentales y Asuntos Constitucionales</i> del Parlamento Europeo	77
3.4.7.4. El <i>Dictamen del Comité Económico y Social Europeo</i>	80
3.4.8. Los contraargumentos	81
3.4.8.1. La confusión con el ser humano	81
3.4.8.2. La confusión con las personas jurídicas. Referencia al esquema societario como posible cobertura para una personalidad jurídica independiente de los sistemas de inteligencia artificial.....	81
3.4.8.3. Juridificación de la ciencia ficción o regulación previsor..	84
3.4.8.4. Futuro y singularidad.....	85
3.4.8.5. Tras la barrera humanista: el dataísmo	87
4. BASES PARA UNA REGULACIÓN	91
4.1. Cuándo	91
4.2. Dónde	92
4.3. Cómo	92
4.3.1. Los derechos afectados y los principios bioéticos	92
4.3.1.1. No maleficencia. No dañar	93
4.3.1.2. Beneficencia. Procurar el bien	95
4.3.1.3. Autonomía. Consentimiento informado	98
4.3.1.4. Justicia. Acceso igualitario a los beneficios de la robótica .	103
4.3.1.5. Precaución. Controlar el riesgo.....	105

4.3.1.6. Transparencia. Contra la opacidad del algoritmo	107
4.3.2. Leyes de Asimov para humanos o derecho natural para robots	108
4.3.3. Contenido de la personalidad algorítmica y requisitos constitutivos	110
Bibliografía	117

Presentación

Javier González sabe liarse la manta a la cabeza y acometer a fondo un tema especialmente complejo que, según él, viene «a propósito de la personalidad jurídica de la inteligencia artificial».

Merece la pena leer sus reflexiones, porque el concepto de persona resulta, como otros conceptos básicos, mucho menos rígido, más rico, más potente. El autor dedica un especial cuidado —incluso amoroso— a las personas electrónicas y, en general, al mundo de la robótica, y conviene prestarle una buena atención, porque la inteligencia artificial y sus efectos y consecuencias van a estar en el centro de todos los debates y también de nuestras inquietudes como seres humanos, que podrían verse superados en cuanto a capacidad mental, o incluso en el terreno de las emociones.

Es un serio trabajo de investigación jurídica que merece también un elogio serio y la petición de que no lo abandone. El paso de la persona física a la jurídica y a la electrónica puede tener todavía más recorrido, y será algo bueno que empecemos a prepararnos sin reserva ni miedos. El ser humano es muy resiliente, y acabará controlando, con mayor o menor prontitud, todos los avances. Eso es lo que ha sucedido hasta ahora. Y si no es así, pues ¡que trabajen ellos!

ANTONIO GARRIGUES WALKER
Diciembre de 2019

Prólogo

La monografía *De la persona a la personalidad algorítmica: a propósito de la personalidad jurídica de la inteligencia artificial*, que tengo el placer de prologar, es un texto de singular interés, tanto para el lector especializado como para aquel que quiera iniciarse en el tema. La obra analiza una de las cuestiones transversales de nuestro tiempo, el lugar que ocupan los algoritmos, qué construcción jurídica merecen como paso previo a una posible regulación y, por ello, el valor que le damos a la tecnología como sociedad. El trabajo incluye una propuesta que combina la aplicación de los clásicos principios bioéticos con el principio de precaución y la exigencia de transparencia para desarrollar una inteligencia artificial en la que confiar.

Su autor, Javier González Granado, posee una interesante doble formación en derecho positivo y bioética. En su condición de jurista, es notario, fue también abogado y adquirió su formación en bioética en el máster en Bioética y Derecho de la Universidad de Barcelona.¹ Fue en su etapa de alumno del máster cuando lo conocí, y tuve la oportunidad de participar en la brillante defensa pública de su tesina final, embrión de la obra que el lector tiene en sus manos, y que, por su calidad y pertinencia, ahora publicamos en la Colección de Bioética de la Cátedra Unesco de Bioética de la Universidad de Barcelona.²

¹ El máster en Bioética y Derecho se imparte en la Universidad de Barcelona desde 1994 en formato semipresencial, y a distancia coordinado por el Observatorio de Bioética y Derecho – Cátedra Unesco de Bioética de la Universidad de Barcelona, centro de investigación interdisciplinar en bioética y derechos humanos (www.bioeticayderecho.ub.edu/es/master).

² La Colección de Bioética del Observatorio de Bioética y Derecho de la Universidad de Barcelona, dirigida por la Dra. María Casado, promueve una concepción de la bioética flexible, pluridisciplinar y laica, en el marco del respeto a los derechos humanos reconocidos, y fomenta el debate informado sobre «las cuestiones éticas relacionadas con la medicina, las ciencias de la vida y las tecnologías conexas aplicadas a los seres humanos, teniendo en cuenta sus dimensiones sociales, jurídicas y ambientales» (art. 1.1., Declaración Universal sobre Bioética y Derechos Humanos, Unesco, 2005). En 2018, la colección recibió el Sello de Calidad en Edición Académica, otorgado por la ANECA y la FECYT, distinción a iniciativa de la Unión de Editoriales Universitarias Españolas (UNE) para reconocer la excelencia científica del proceso editorial de las colecciones publicadas por universidades y editoriales científicas de carácter privado (www.bioeticayderecho.ub.edu/es/libros).

La regulación de la inteligencia artificial es prioritaria para las agendas gubernamentales y, además, desde el punto de vista de su aplicación, inunda todas las esferas de nuestra vida, hasta afectar a nuestra propia identidad. Pegados al móvil y a tabletas como vivimos, casi en simbiosis, estamos confiriendo al algoritmo un lugar preeminente; situación sobre la que ya se están llevando a cabo interesantes reflexiones y sobre cuyo impacto en la vida de los ciudadanos es urgente tomar conciencia. Los dispositivos digitales integran cierta inteligencia artificial incrustada (como es el caso de los asistentes virtuales Siri, de Apple, o Cortana, de Microsoft). La inteligencia artificial se nutre de conjuntos de datos, incluidos los de carácter personal, que alimentan a los algoritmos: operaciones matemáticas complejas que siempre han existido y que ahora viven su apogeo debido al desarrollo de la ciencia y de la técnica, y gracias a la ingente cantidad de datos de la que disponemos. En definitiva, la inteligencia artificial permite mejorar la toma de decisiones mediante predicciones a través de algoritmos, superando las capacidades humanas. No ha venido para sustituirnos, sino para ayudarnos. De datos y algoritmos, y de humanos que deben corregir los sesgos que estos últimos incorporan, va nuestro tiempo. El algoritmo condiciona así nuestro desarrollo como individuos y como especie humana.

Estas son las cuestiones de las que trata la obra de Javier González Grando. El trabajo contribuye a dibujar la respuesta jurídica que permita regular la relación entre inteligencia humana y artificial. La sociedad digital guiada por el dato requiere este tipo de aportaciones que combinan la perspectiva ética con el análisis jurídico. El mercado único digital, propiciado por la convergencia de tecnologías, permitirá a Europa competir a escala global y participar en la *data política* para dar la batalla a Estados Unidos y China. El mérito consiste en ser, entonces, de los primeros en estudiar esos territorios poco explorados, como el estatuto jurídico del algoritmo, para suscitar la reflexión y el pensamiento crítico.

Investigar sobre las cuestiones de nuestro tiempo, preparar un manuscrito, las revisiones hasta publicar el libro ¡no es trivial! Tampoco es baladí publicar con el sello Unesco, que en el ámbito de la educación superior supone uno de los máximos reconocimientos a los que se puede aspirar y que proyecta internacionalmente una trayectoria fundamentada en la investigación, la educación y la transferencia de conocimiento, sellos distintivos del Observatorio de Bioética y Derecho de la Universidad de Barcelona.³ En particular, destaca la trans-

³ El Observatorio de Bioética y Derecho – Cátedra Unesco de Bioética de la Universidad de Barcelona se creó en el año 1995 para analizar interdisciplinariamente las implicaciones éticas, jurídicas y sociales de la biomedicina y la biotecnología (www.bioeticayderecho.ub.edu/es/presentacion).

ferencia del conocimiento generado para aportar los argumentos necesarios para fomentar un debate social informado sobre las cuestiones bioéticas. El objetivo final es contribuir a la creación de una sociedad más transparente y democrática y a la toma de decisiones libre e informada.

Obras como la de Javier González Granado permitirán impactar en el legislador y en los decisores políticos, pero también serán de utilidad para periodistas y comunicadores científicos como canal de transmisión fundamental en la sociedad para aumentar la calidad del debate. Atendiendo a su blog⁴ y a sus intervenciones en los medios, el autor no es nuevo en esto. Así, serán diversos los perfiles de los lectores que se van a interesar por esta obra, porque es pertinente, necesaria, oportuna y, además, permite invitar al debate sobre este asunto tan crucial de nuestro tiempo a aquellos perfiles más técnicos, como el científico de los datos, el ingeniero, etc. Para reflexionar de manera conjunta sobre cómo regular la vida regida por complejas operaciones matemáticas; sobre el lugar que le damos a los datos y a los algoritmos en el siglo XXI, en el que los avances científicos y tecnológicos carecen de una ética adaptada a su esencia y potencialidad.

Estamos acostumbrados a leer y a ver en los medios de comunicación ejemplos de iniciativas para desarrollar algoritmos que han sido abandonadas, porque las decisiones resultantes, que pretendían mejorar las tomadas por los humanos, eran discriminatorias. Los sesgos que incorpora la inteligencia artificial deben ser corregidos para que no se perpetúen. Los sistemas de aprendizaje profundo, que pueden llegar a tomar decisiones por sí mismos, deberían contar con datos de calidad y que, además, no incorporen estos sesgos. El caso de la inteligencia artificial de Amazon para seleccionar al mejor candidato fue sonado. El propio Jeff Bezos, dueño de la compañía, anunció que no ofrecería este servicio porque sistemáticamente el algoritmo nunca priorizaba a una mujer como candidata aunque tuviera el mejor currículum.⁵ La realidad alimenta la artificialidad. Lo mismo ocurre con el discurso de odio en las redes sociales. Pensemos también en los sistemas de crédito social basados en la vigilancia intensiva del ciudadano y la constante correlación de todos sus datos.

La Ley Orgánica de Protección de Datos Personales y Garantía de los Derechos Digitales (diciembre de 2018),⁶ en línea con el Reglamento General de

⁴ <https://tallerdederechos.com>.

⁵ Reuters Technology, «Amazon abandona un proyecto de IA para la contratación por su sesgo sexista», 14 de octubre de 2018 (<https://es.reuters.com/article/technology/idESKCN1MOoM4>).

⁶ Ley Orgánica 3/2018 de 5 de diciembre de Protección de Datos Personales y garantía de los derechos digitales. «BOE» núm. 294, de 06/12/2018 (www.boe.es/buscar/act.php?id=BOE-A-2018-16673).

Protección de Datos (de aplicación desde mayo de 2018),⁷ incluye derechos que se refieren a la alfabetización digital. Como ciudadanos, no podemos prescindir de estos conocimientos, de esta formación sobre digitalización, porque de otra forma seremos más esclavos que libres y se tomarán decisiones por nosotros creyendo que somos nosotros los que las estamos tomando, lo que aumentará todavía más las desigualdades, pero de forma encubierta.

Recientemente, los científicos de los datos reclaman formación en ética, en intimidad y protección de datos, clases que nunca recibieron y que ahora son más necesarias que nunca, puesto que de lo que se trata es de tomar decisiones equitativas y determinar la responsabilidad sobre el algoritmo. Por ello, desde la bioética, como soporte al derecho, y también como paso previo a la opción política, es necesario reflexionar sobre la regulación jurídica del algoritmo.

Se hace preciso, entonces, entender los retos a los que nos enfrentamos como sociedad y definir los conceptos que están en juego para un adecuado tratamiento jurídico acompañado de la reflexión ética. Para entender la inteligencia artificial y reflexionar desde la bioética o la *biodataética*, es necesario entender que nuestros datos, nuestra intimidad, son relevantes. Debido a las posibilidades tecnológicas, al conocimiento científico disponible y a la decisión política de crear bases de datos de carácter personal de distinta índole para, más tarde, correlacionarlas, reutilizando los datos disponibles, hemos dejado de ser anónimos para ser reidentificables. El quid de la cuestión radica, entonces, en proteger la intimidad a través de la protección de los conjuntos de datos personales que, veinticuatro horas siete días a la semana se emiten, almacenan y tratan en la sociedad digital. Es esencial garantizar la *seudonimización*⁸ para asegurar que el desarrollo de algoritmos se efectúa sin posibilidad de atribuir personalidad a los conjuntos de datos de los que bebe.

Somos conjuntos de datos que, precisamente, por no ser nosotros ni famosos ni relevantes se convierten en objeto de deseo, porque contribuyen a crear ontologías que permiten desarrollar las inteligencias artificiales fundadas en

⁷ Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos) (Texto pertinente a efectos del EEE). OJ L 119, 4.5.2016, págs. 1-88 ELI: <http://data.europa.eu/eli/reg/2016/679/oj> (EURLEX).

⁸ La seudonimización como técnica se refiere al procesamiento de datos personales de tal forma que estos no puedan atribuirse a una persona sin el uso de información adicional, siempre que esta última se mantenga de forma separada y esté sujeta a medidas técnicas y organizativas para asegurar que los datos personales no se atribuyen a una persona identificada o identificable. European Commission, How to complete your Ethics Self Assessment, 4 de febrero de 2019 (https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-self-assess_en.pdf).

algoritmos con distintos fines. Abrir el móvil utilizando datos biométricos con la cara o la huella de una persona —una situación que parece una experiencia única— resulta que permite, a la vez —y para ello se habilita—, nutrir la base de datos de Android o de Apple para perfeccionar su inteligencia artificial, y la de Microsoft y Google, por no citar a los gigantes asiáticos. Amplía, también, los posibles mercados de datos basados en la monetización de perfiles.⁹ Lo que está en juego, y es el desafío que inunda toda la obra de Javier González Granado, es cómo proteger la intimidad de las personas humanas a la vez que conferir un estatus jurídico adecuado a la personalidad algorítmica.

La regulación de la inteligencia artificial cuestiona el valor que tradicionalmente le hemos dado a los datos personales y a la intimidad. Un mercado digital único europeo se constituye sobre bases de datos en las que conjuntos de datos, incluidos los de carácter personal, se explotan en aras de intereses personales e iniciativas orientadas al bien común. Así, la materia prima de la investigación, el desarrollo y la innovación de la inteligencia artificial son este tipo de datos.

Si nos centramos en el ámbito biomédico, las investigaciones se basan en la convergencia de tecnologías, como la aplicación de la analítica de datos masivos (*big data*) y la inteligencia artificial. De hecho, si revisáramos la hemeroteca digital, veríamos cómo en poco tiempo se ha dejado de hablar del *big data* para darle todo el protagonismo a la inteligencia artificial, pero ambos caminan de la mano e implican correlacionar datos de todo tipo —genéticos, biométricos, etc.—. La expresión *big data* se refiere «al tratamiento de grandes volúmenes de datos mediante algoritmos matemáticos con el fin de establecer correlaciones entre ellos, predecir tendencias y tomar decisiones».¹⁰ Es comprensible, entonces, que el interés actual por el acceso a historias clínicas almacenadas en bases de datos digitales sea creciente, pues en estas se encuentran conjuntos de datos personales bien estructurados, bajo criterios de calidad, fiabilidad y seguridad. Además, no solo contienen categorías especiales de datos, como los datos de salud, genéticos, etc., sino también sociodemográficos. El acceso a la información almacenada en estas bases de datos permite conocer también el código postal, la fecha de nacimiento y el sexo, y solo con estos tres

⁹ En este sentido, consúltese la obra colectiva CASADO, M. (coord.) (2016). *De la solidaridad al mercado*. Fontamara. Disponible en abierto en www.bioeticayderecho.ub.edu en su reedición a cargo de Edicions de la Universitat de Barcelona en 2017.

¹⁰ LLÀCER, M. R.; CASADO, M. y BUISÁN, L. (2015). *Documento sobre bioética y Big Data en salud: explotación y comercialización de los datos de los usuarios de la sanidad pública*. Barcelona: Publicacions i Edicions de la Universitat de Barcelona, pág. 33. Disponible en: <http://hdl.handle.net/2445/104585>.

datos es posible reidentificar a las personas en un porcentaje muy elevado.¹¹ Asimismo, en el sistema sanitario público, la historia clínica recoge aspectos de nuestra intimidad muy delicados, como son nuestro consumo farmacéutico o los hijos que están a nuestro cargo en la cotización de la seguridad social. A diferencia de otros datos, de dudosa procedencia y fiabilidad que podríamos encontrar en otras bases de datos en la red sobre nosotros mismos, la historia clínica es unpreciado y codiciado objeto para la iniciativa pública y privada.

La inteligencia artificial en el ámbito de la salud sirve para desarrollar una medicina personalizada o predecir los efectos adversos de los medicamentos en menos tiempo. Lo que el humano tardaría años en hacer, la inteligencia artificial lo efectúa en segundos; nos ayuda, entonces, sobremanera, pero nunca puede, ni debe, sustituirnos. Es habitual ver anuncios de *Hackatones* o retos para desarrollar algoritmos en el ámbito médico como parte de proyectos de investigación para, por ejemplo, predecir el riesgo a desarrollar determinada patología. La forma en la que se hace investigación en el siglo xxi es bien distinta a la del siglo xx, en la que el paradigma era el ensayo clínico promovido, en buena parte, por la industria farmacéutica. Hoy son diversos los terceros con intereses en el terreno de la investigación e innovación biomédica basada en *big data* e inteligencia artificial. De hecho, son varios los sistemas de inteligencia artificial que, de forma rutinaria, explotan datos y aprenden, e interpretan, el contenido de las historias clínicas para mejorar la toma de decisiones. Para ello, cedemos buena parte de nuestra intimidad por razones solidarias, para recibir en el futuro intervenciones y tratamientos. Somos, entonces, titulares y usuarios finales de la materia prima, el oro de nuestro tiempo.

Esta nueva realidad ha cambiado también la propia esencia del consentimiento informado. El mundo hiperconectado, que tiende a la acumulación de datos, modifica por completo las reglas del juego, que siempre se han sustentado en un consentimiento explícito para un acto concreto después de recibir información adecuada —con excepciones, tasadas—. Ahora, por el contrario, transitamos hacia modelos de consentimiento presunto; esto es, todos estamos en la base de datos por defecto. Esto no significa que ambos modelos de consentimiento no sean compatibles. Lo que quiero resaltar es que la aplicación de las antiguas reglas para evaluar, por ejemplo, la pertinencia de una investigación en personas con datos personales o muestras biológicas proveniente del siglo xx deviene ahora ineficaz si no comprendemos las condiciones en las que

¹¹ SWEENEY, L. «Simple demographics often identify people uniquely». Carnegie Mellon University, editor. Data Privacy Working Paper 3. 2000.

el algoritmo se desarrolla y su propia esencia. Desde la reflexión bioética, conviene potenciar la aplicación de los principios de transparencia y rendición de cuentas para compensar las posibles carencias que puedan provocar modelos de consentimiento presunto.¹²

Otro ejemplo ilustrativo lo proporciona la existencia de dispositivos de salud que envían alertas para informar o anticipar posibles situaciones críticas para un ciudadano/paciente. Aplicaciones y dispositivos directamente disponibles para el consumidor, para los que se reclaman sistemas de certificación estandarizados que garanticen que sus creadores merecen la confianza de los usuarios finales (que recuerden, como se ha indicado antes, que se nutren de los datos de estos).¹³ El sello de calidad de los algoritmos, que depende en buena parte de la reputación de las entidades en las que estos se prueban y comprueban, y de la calidad de los datos que manejen, está por llegar. Son numerosas las iniciativas que llaman a las puertas de grandes hospitales y centros de investigación de referencia para obtener el aval ético de sus fórmulas algorítmicas mediante procesos de investigación y pruebas de concepto de que deben ser evaluadas por los comités de ética de la investigación acreditados. No entraré a analizar, como he hecho en otras ocasiones, la falta de formación de los comités de ética de la investigación en estos aspectos, y la necesidad de integrar miembros expertos en estas áreas para evaluar adecuadamente los proyectos que se proponen.¹⁴

Para situar la aportación de Javier González Granado, conviene revisar los pilares sobre los que se sustenta la inteligencia artificial europea. Desde la perspectiva de la protección de datos, y para asegurar que los ciudadanos retomen el control sobre los datos de los que son titulares, el Reglamento General de Protección de Datos asienta de forma transversal los principios y los derechos que se deben aplicar y respetar en el mundo algorítmico. Los tratamientos de los datos personales para desarrollar algoritmos deben basarse en los siguientes principios: licitud, lealtad y transparencia; limitación del propósito; minimiza-

¹² Sobre este tipo de modelos de consentimiento presunto para utilizar bases de datos de salud públicas con fines de investigación, consúltese el capítulo DE LECUONA, I. y VILLALOBOS, M. J. «El valor y el precio de los datos personales de salud en la sociedad digital», en GARCÍA MANRIQUE, R. (coord.) (2018). *El cuerpo diseminado: Estatuto, uso y disposición de los biomateriales humanos*. Thomson Reuters Aranzadi, Cizur Menor, págs. 171-189.

¹³ En este sentido, véase, por ejemplo, el proyecto de investigación NESTORE: financiado por el Programa Marco de Investigación Horizonte 2020 de la Unión Europea para desarrollar un sistema de consejo virtual para un envejecimiento activo. Cordis: <https://cordis.europa.eu/project/id/769643/es>.

¹⁴ DE LECUONA, I., «Evaluación de los aspectos metodológicos, éticos, legales y sociales de proyectos de investigación en salud con datos masivos (big data)», *Gaceta Sanitaria*, vol. 32, núm. 6, págs. 576-578 (noviembre-diciembre de 2018) DOI: 10.1016/j.gaceta.2018.02.007.

ción del dato; exactitud y actualización; almacenamiento limitado; integridad y confidencialidad, y rendición de cuentas.

Asimismo, deben aplicarse por defecto los principios de privacidad desde el diseño y la privacidad para determinar las medidas técnicas y organizativas necesarias para asegurar la protección de datos de carácter personal, desde el diseño de la intervención que se propone y durante su desarrollo. También deben llevarse a cabo, en función de la tipología de datos a tratar, las correspondientes metodologías para la evaluación del impacto de los tratamientos de datos propuestos en las personas afectadas. Es el caso de las categorías especiales de datos, como, por ejemplo, los datos de salud.¹⁵ El Reglamento introduce, además, el Delegado de Protección de Datos como figura independiente para asesorar en estos procesos.

En cuanto a los derechos a ejercitar frente a la programación y la aplicación de algoritmos: derecho a ser informado, derecho de acceso, derecho de rectificación, derecho al borrado/olvido, derecho a restringir el procesamiento de los datos, derecho a portabilidad de los datos y derecho a no ser objeto de una decisión automatizada, que incluye la controvertida elaboración de perfiles, decisiones que deben incorporar la intervención/corrección humana. Se amplía así el catálogo de derechos reconocidos para la protección de la intimidad y la confidencialidad a la estela de los conocidos como derechos ARCO (acceso, rectificación, cancelación y oposición), en los que se asienta la autodeterminación informativa, reconocida por tribunales en el siglo xx.¹⁶ De la imprenta al CD y del USB a la nube, ¡el cambio es radical! El Reglamento introduce un régimen proactivo frente a la antigua normativa de carácter reactivo. Además, la información debe transmitirse de forma clara y simple para que sea inteligible para el destinatario, y en el caso de que se produzca una brecha de seguridad, se debe notificar en menos de setenta y dos horas a la autoridad de protección de datos correspondiente. Las multas pueden ser de hasta el 4 % de los beneficios globales anuales para las empresas. Una cifra nada despreciable, además de las nefastas consecuencias que el incumplimiento pueda tener en la reputación de los infractores.

En cuanto a la gestión del ciclo de vida de los datos que alimentan los algoritmos y, como mínimo, en el ámbito de los proyectos de investigación financiados por fondos públicos europeos, se deben aplicar los principios conocidos

¹⁵ Agencia Española de Protección de Datos. *Listado de tipos de tratamientos de datos que requieren evaluación del impacto relativa a la protección de datos* (www.aepd.es/sites/default/files/2019-09/listas-dpia-es-35-4.pdf).

¹⁶ En este sentido, se recomienda la lectura del libro de SÁNCHEZ URRUTIA, A. y NAVARRO MICHEL, M. (2003). *Tecnología, intimidad y sociedad democrática*. Barcelona: Icaria.

como FAIR (por sus siglas en inglés).¹⁷ Así, los datos deben estar disponibles y ser accesibles, interoperables y reutilizables. Al principio de calidad de los datos se le suma la interoperabilidad para un uso adecuado de estos que permita así su legibilidad e interpretación de forma estandarizada. La reutilización se impone vía Directiva relativa a los datos abiertos y la reutilización de la información del sector público.¹⁸ Conviene recordar aquí que Europa apuesta por un modelo de ciencia abierta que promueve poner a disposición los datos generados en la investigación y que rompe también los tradicionales esquemas de la generación de conocimiento propios del modelo analógico.¹⁹

En marzo de 2019 participé en la presentación pública de la Estrategia Española de Investigación, Desarrollo e Innovación²⁰ para tratar sus aspectos éticos considerados como una prioridad. Según la Estrategia, la inteligencia artificial debe evitar el sesgo negativo, los prejuicios de género y cualquier forma de discriminación. En su diseño, es necesario integrar un enfoque interdisciplinar que incorpore los aspectos éticos, legales y sociales. Se trata de alinear los intereses sociales y los objetivos de los distintos actores implicados. Destaca, asimismo, la necesaria elaboración de un código de ética en inteligencia artificial y refiere a la *Declaración de Barcelona*,²¹ que defiende una aplicación de la inteligencia artificial prudente, confiable, responsable, identificable y con una autonomía regida por un conjunto de reglas que velen por el acervo humano.

Desde el análisis bioético, son numerosas las aportaciones de organizaciones internacionales de referencia: del Consejo de Europa con la *Guía sobre Inteligencia Artificial y Protección de Datos* (2019)²² y la *Guía para la protección*

¹⁷ European Commission Expert Group on FAIR Data, Turning FAIR into reality (2018). *Final Report and Action Plan* (https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf).

¹⁸ Directiva (UE) 2019/1024 del Parlamento Europeo y del Consejo, de 20 de junio de 2019, relativa a los datos abiertos y la reutilización de la información del sector público.

¹⁹ Para más información sobre el modelo de ciencia abierta, véase la Iniciativa de Datos abiertos del Gobierno de España (<https://datos.gob.es/es/noticia/principios-fair-buenas-practicas-para-la-gestion-y-administracion-de-datos-cientificos>).

²⁰ Ministerio de Ciencia e Innovación (2019). *Estrategia Nacional de Investigación, innovación y desarrollo en Inteligencia Artificial*. Madrid (www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_IDI.pdf).

²¹ *Declaración de Barcelona* para el desarrollo y uso adecuado de la inteligencia artificial en Europa, Barcelona, 2017, promovida principalmente por el Dr. Ramón López de Mántaras, director del Instituto de Investigación en Inteligencia Artificial del CSIC, y Luc Steels, investigador ICREA del Instituto de Biología Evolutiva de la Universidad Pompeu Fabra (www.iii.csic.es/barcelonadeclaracion/).

²² Council of Europe (2019). *Guidelines on Artificial Intelligence and Data Protection*. Estrasburgo (www.coe.int/en/web/artificial-intelligence/-/new-guidelines-on-artificial-intelligence-and-data-protection).

de las personas con respecto al tratamiento de datos de carácter personal en el mundo *Big Data* (2017);²³ de la Unesco también sobre los usos del *big data* (2017)²⁴ y los trabajos de la Comisión Europea a través del Grupo Europeo de Ética y Nuevas Tecnologías, informe la Ética de la Inteligencia Artificial (2018),²⁵ que han sembrado la semilla de opiniones futuras por parte de grupos *ad hoc* y especializados. En este sentido, destaca la contribución del Grupo de Expertos de Alto Nivel para la Inteligencia Artificial que, en 2019, publicó una guía para una inteligencia artificial confiable y también una metodología fundamentada en la Carta Europea de Derechos Fundamentales para que las organizaciones públicas y privadas, incluyendo también a la pequeña y mediana empresa, puedan evaluar el estado de la cuestión.²⁶ La estrategia europea, recién creada (febrero de 2020), se centra así en una inteligencia artificial confiable,²⁷ tal y como propugna el citado grupo, y que debe seguir los siguientes principios: asegurar una inteligencia artificial centrada en el humano, fundamentada en el respeto por los derechos humanos y los valores a respetar en Europa; la beneficencia y no maleficencia, el respeto por la autonomía de los humanos, la justicia y la explicabilidad. Se trata de evaluar los posibles efectos de la inteligencia artificial en los humanos y en el bien común. Destaca la preocupación del Grupo de Alto Nivel por las asimetrías en cuanto a la información de la que puedan disponer los distintos actores, por ejemplo, empleadores y empleados, etc., y reclama atención hacia aquellas

²³ Council of Europe. Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Convention 108) (2017). *Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data*, Estrasburgo (<https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806ebe7a>).

²⁴ International Bioethics Committee (2017). *Report on Big Data and Health*. París: Unesco (<https://unesdoc.unesco.org/ark:/48223/pf0000248724>).

²⁵ European Group on Ethics of Science and New Technologies (2018). *Statement on artificial intelligence, robotics and 'autonomous' systems* (https://ec.europa.eu/info/publications/lege-statements_en). La traducción al castellano por el Observatorio de Bioética y Derecho de la UB está disponible en www.bioeticayderecho.ub.edu/es/el-obd-traduce-los-informes-del-grupo-europeo-de-etica-de-la-ciencia-y-las-nuevas-tecnologias.

²⁶ High level Expert Group on Artificial Intelligence (2019). *Ethics Guideline on Artificial Intelligence, European Commission* (<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>).

²⁷ *Communication to the Commission. European Commission Digital Strategy. A digitally transformed, user-focused and data-driven Commission*, Bruselas, 21.11.2018 C(2018) 7118 final (https://ec.europa.eu/info/sites/info/files/strategy/decision-making_process/documents/ec_digitalstrategy_en.pdf) y *White Paper on Artificial Intelligence: A European approach to excellence and trust*, Bruselas, 19.2.2020 COM(2020) 65 final (https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf).

situaciones que puedan comprometer los derechos de los colectivos y personas más vulnerables.

Son requisitos para una inteligencia artificial confiable: la rendición de cuentas, el diseño para todos, la gobernanza del dato, la supervisión y el control humano, la exigencia de no discriminación y garantizar la seguridad y la transparencia, de tal forma que sea posible el desarrollo de la inteligencia artificial robusta, segura y respetuosa con la intimidad y el tratamiento confidencial de los datos personales. Según el Grupo de Alto Nivel, estas cuestiones no pueden reducirse simplemente a marcar casillas; una inteligencia artificial confiable requiere un exhaustivo análisis, desde el diseño de la intervención y un seguimiento continuo, puesto que el algoritmo requerirá correcciones y adaptaciones guiadas por las decisiones humanas. La apuesta por una Europa líder global en innovación es clara, pero no a costa de los derechos de sus ciudadanos, sino asegurando su bienestar y desde esta orientación basada en el respeto de los derechos fundamentales.

No hace mucho, la canciller alemana Angela Merkel manifestaba su preocupación por la dependencia excesiva que Europa tiene del conocido como imperio GAFAM (Google, Amazon, Facebook y Microsoft) de origen estadounidense, reclamando infraestructuras digitales propiamente europeas.²⁸ Mariana Mazzucato,²⁹ economista italiana y profesora del University College London, alerta sobre la necesidad de prevenir el feudalismo digital.³⁰ Para ello, anima a establecer mecanismos de cocreación de valor sobre los datos como antídoto a la salvaje extracción de valor imperante.

El Libro Blanco de la Inteligencia Artificial europea, para cuya elaboración el nuevo gabinete europeo tan solo ha tenido cien días, se refiere a la creación de inteligencias artificiales y de infraestructuras digitales al servicio de las personas y desde el respeto por sus derechos. En la misma línea que establece Javier González Granado hacia el final de su obra, Europa apuesta, como hemos visto, por una inteligencia artificial ética y confiable.

Escribo este prólogo mientras leo las últimas páginas del fabuloso libro *Máquinas como yo*, de Ian McEwan (Anagrama, 2019). Situada su acción en la década de 1980 y cambiado el final de algunos acontecimientos históricos,

²⁸ *El País*, «Europa ultima un plan para dar la batalla de los datos», domingo, 24 de noviembre de 2019 (https://elpais.com/economia/2019/11/16/actualidad/1573926886_318836.html).

²⁹ Autora de *El estado emprendedor: Mitos del sector público frente al privado*, Barcelona: RBA Economía y El valor de las cosas. Barcelona: Taurus, 2019.

³⁰ *El País*, «Prevenir el feudalismo digital», 13 de octubre de 2019 (https://elpais.com/elpais/2019/10/11/opinion/1570808221_636252.html).

como, por ejemplo, que Alan Turing no se haya suicidado, permite al protagonista humano tener una conversación sobre la inteligencia artificial llamada Adán, que ha comprado y luego programado a medias con su novia y vecina, y que, aparentemente, carece de moral. Imagino a Alan Turing conversando con Javier González Granado sobre cómo regular inteligencias artificiales que llegan a tomar decisiones sobre su propio destino. En definitiva, ¿dónde están los límites éticos y jurídicos de la inteligencia artificial? Les aconsejo que lean el libro de Javier González Granado antes de responder a la pregunta.

ITZIAR DE LECUONA
Barcelona, febrero de 2020

Introducción

El 31 de mayo de 2016, la Comisión de Asuntos Jurídicos del Parlamento Europeo aprobó el *Proyecto de informe con recomendaciones a la Comisión sobre normas de Derecho Civil sobre robótica*, en el que establecía las directrices de la futura regulación europea en materia de robótica e inteligencia artificial.

El documento expone los principales riesgos y desafíos que plantean estas innovaciones tecnológicas, hace un recorrido por los derechos más destacados que pueden verse afectados y propone diversas soluciones jurídicas y un marco ético de principios a respetar de honda raigambre bioética. Dentro de estas soluciones, la más llamativa, por su originalidad y trascendencia, es la propuesta de crear un nuevo tipo de personalidad jurídica a la que califica como «personalidad jurídica electrónica», y que se concedería a aquellos sistemas de inteligencia artificial que sean complejos y que puedan tomar decisiones autónomas inteligentes o interactuar con terceros de forma independiente.

Este planteamiento, confirmado en la Resolución definitivamente aprobada por el Parlamento Europeo el 16 de febrero de 2017, se presenta como una posible solución al problema de deslindar responsabilidades en caso de daños ocasionados por tales sistemas autónomos pues, según tales documentos, la normativa tradicional no bastará para establecer su responsabilidad, ni en el ámbito contractual ni en el de la responsabilidad extra-contractual.

Entre ambos documentos se produjeron destacadas reacciones, doctrinales, sociales e institucionales, contrarias a tal solución, argumentando que supondría una total vulneración de los principios humanistas sobre los que descansa nuestra civilización. Sin duda, tales opiniones incidieron en el Parlamento Europeo, pues la Resolución ofrece en relación con el reconocimiento de esa eventual personalidad electrónica una solución mucho más matizada que la Propuesta de informe. Aun hablando expresamente de esta categoría de personas electrónicas responsables de reparar los daños, lo hace remitiendo esa regulación al indeterminado momento del «largo plazo».

La finalidad de este trabajo es investigar, desde una perspectiva jurídica, la solución propuesta por ambos documentos de la Unión Europea. Para ello, y con carácter previo, se hace un recorrido histórico por el concepto de persona,

partiendo del momento anterior a la aparición del Derecho como mecanismo regulador de conductas humanas, y, por lo tanto, anterior a su surgimiento como elemento subjetivo de relaciones jurídicas, hasta la concreción de aquel concepto y su equiparación con el ser humano. Se hará referencia a las opiniones doctrinales más destacadas sobre el particular, para construir sobre estas varios modelos esquemáticos basados en la combinación de los tres elementos que encarnan, como esenciales, las principales concepciones de la persona antropológica, filosófica y jurídica: pertenencia a la especie humana, facultad de autoconciencia y funcionalidad jurídica.

A continuación, se exponen las razones manejadas por los autores para fundamentar la existencia de la comunidad moral que ampara a todos los seres humanos y las desviaciones que se han producido en los ordenamientos jurídicos por defecto o por exceso, encajando esos supuestos extravagantes en alguno de los modelos propuestos.

El análisis concreto de la personalidad electrónica o algorítmica parte de una breve referencia histórica, deslindando leyenda y realidad y poniendo de relieve la importancia del desarrollo de estos sectores tecnológicos hasta el punto de que reclaman la aparición de disciplinas específicas en los ámbitos ético y jurídico. Y, desde esa base, se perfila la posible conceptualización de los sistemas de inteligencia artificial como agentes morales, los requisitos que determinan este concepto y los cambios que en su concepción tradicional han realizado algunos estudiosos para incluir su encaje en él.

Tras la exposición resumida del Proyecto de informe y de la Resolución del Parlamento Europeo, se contraponen los principales argumentos contrarios y favorables a la personificación de los sistemas de inteligencia artificial, con una referencia en especial a la posibilidad de utilizar, sin necesidad de ningún cambio legislativo, los esquemas tradicionales de las personas jurídicas para dar acomodo a una personalidad jurídica de un sistema de inteligencia artificial independiente de cualquier persona natural.

De esta forma, se exponen algunas estructuras jurídicas monstruosas y paradójicas, como sociedades sin socios y/o órgano de administración, y se analiza la posibilidad de su subsistencia jurídica en los ordenamientos jurídicos de Alemania, Suiza, Reino Unido y Estados Unidos, siguiendo las principales líneas de investigación jurídica existentes. La cuestión permanecía inédita en el Derecho español, y en este trabajo se incluye la primera aproximación para nuestro ordenamiento jurídico.

La exposición de los argumentos favorables a la personalidad jurídica de los sistemas de inteligencia artificial concluye con una referencia a la concepción de la persona bajo una óptica dataísta.

El trabajo finaliza con una propuesta de bases para una regulación normativa, partiendo de los principios bioéticos y jurídicos proclamados en los documentos de la Unión Europea y perfilando la estructura, el contenido y el funcionamiento de esta eventual y novedosa personalidad algorítmica.

En una primera aproximación al concepto de persona, se afirma que *persona*, *sujeto de derecho* y *ser humano* son sinónimos, y esta acepción vulgar tiene también su reflejo en el ámbito filosófico y jurídico. Así, la *Declaración Universal de los Derechos Humanos* vincula este concepto a la pertenencia a la familia humana, y le atribuye un sentido relacional al vincularlo a la dignidad intrínseca al ser humano en la que se basan «la libertad, la justicia y la paz en el mundo».

Desde una perspectiva jurídica, la persona es el elemento subjetivo del Derecho, es el eje alrededor del cual gira el Derecho, en el sentido de que todo el Derecho existe en función de las personas (Castiella, 2003: 21); paralelamente, la personalidad es la aptitud, reconocida por el ordenamiento jurídico, para ser titular de derechos y obligaciones.

Es lugar común en la doctrina jurídica afirmar que en la actualidad esa posición (persona, sujeto de derechos) corresponde al ser humano (Castro, 2008: 30) y, traslaticiamente, a ciertas organizaciones jurídicas. Un recorrido histórico ayuda a determinar qué hay de acertado en esa visión, si la persona jurídica lo es solo «traslaticiamente» y si en realidad son sinónimos los términos «ser humano», «persona» y «sujeto de derechos».

1.1. *HOMO SAPIENS, HOMO JURIDICUS*

1.1.1. Hordas, tribus, jefaturas y Estados

Enlazando con la concepción aristotélica del ser humano como animal político y ser social, los juristas romanos delimitaron con claridad la sociedad humana como el ecosistema en el que nace y se desarrolla el Derecho. Al mismo tiempo, pusieron de relieve el carácter biunívoco de la relación entre ambos elementos: no existe sociedad humana sin Derecho ni es posible la existencia del Derecho sin una sociedad humana que lo sustente.

Orden coactivo y soberano de conducta de carácter externo (a diferencia del carácter interno de las normas morales), el Derecho no surge de forma

automática con la aparición de *Homo sapiens sapiens* hace unos 140 000 años. Nace como un producto cultural vinculado a la desaparición de las sociedades cazadoras-recolectoras y a la jerarquización de las primitivas sociedades agrarias, lo que no acontece antes del año 9000 a. C. en Mesopotamia, fenómeno que más tarde se repite en India (6000 a. C.), China (5000 a. C.), Europa (4500 a. C.), África (3000 a. C.) y América (2500 a. C.).

No obstante lo anterior y una vez explicado el «sentido moral» por referencia a una base genética-evolutiva que permite la deliberación ética (Sánchez y Zavaleta, 2011: 111-135), y admitida la presencia de conductas morales en animales no humanos, es decir, en los primates (Brosnan, 2013: 249-272), debe reconocerse que el proceso de hominización, iniciado unos cuatro millones de años antes de la aparición del Derecho como tal, fue acompañado de una progresiva normativización de las relaciones, a medida que aumentaba su número y complejidad. Y esta normativización progresiva culminó con el tránsito de una regulación exclusivamente moral a una regulación jurídica.

Esta idea enlaza con la distinción propuesta por Michael Gagarin entre sociedades prejurídicas, sociedades protojurídicas y sociedades jurídicas (Gagarin, 1989: 8). También se relaciona con la explicación antropológica de la evolución de las sociedades humanas en hordas, tribus, jefaturas y Estados. Lo cierto es que esta clasificación presenta el inconveniente de lo artificioso y difuso de los límites entre esas categorías, pero se le ha de reconocer la virtud de permitir suplir, con la referencia a ejemplos de pueblos cazadores-recolectores actuales, la falta de fuentes escritas en las primeras etapas de la humanidad.

1.1.2. Sociedades prejurídicas

Puede sostenerse que en las sociedades estructuradas en hordas y tribus (el esquema propio de los pueblos cazadores-recolectores, grupos humanos de decenas o, a lo sumo, centenas de miembros) se tiende

[...] a un equilibrio de poder basado en las relaciones y consideraciones de igualdad, al desarrollo de vínculos afectivos y a mecanismos regulatorios que refuerzan los comportamientos positivos e inhiban o restrinjan los negativos y [...] en estas circunstancias, los comportamientos que son eficaces se abstraen como normas de comportamiento debidas por todos los miembros del grupo. Como estas normas que reflejan el *ethos* colectivo no son producidas intencionadamente con el fin de solucionar los conflictos [...] no se requeriría ni de un pacto social ni de una creación *ex-profeso* de normas jurídicas, por lo que en este tipo de organiza-

ciones sociales no se producen las distinciones conceptuales entre Moral y Derecho (Bernal, 2011: 169).

En estos conjuntos normativos prejurídicos falta una nota esencial de la norma jurídica del Derecho: su heteronomía o exterioridad al sujeto; tampoco es posible localizar en ellos las demás notas que, según Hart (1994: 100-110), caracterizan al Derecho, pues aunque puedan existir las llamadas normas primarias —reguladoras de conducta—, faltan las normas secundarias, las encargadas de regular la creación y aplicación de las normas primarias y, especialmente, la llamada norma de reconocimiento que permite a los sujetos reconocer aquellas normas primarias como reglas de comportamiento.

Esta normativización prejurídica es coherente con un marco socio-económico en el que no existe la propiedad privada, sino una mera apropiación de cosas muebles de uso personal (no transmisibles por herencia, sino que acompañan al difunto en su tumba), en la que el parentesco es el vínculo base de las relaciones (aunque pueda tener un carácter más social que biológico) y en la que los intercambios se basan en el principio de reciprocidad alimenticia no redistributiva (por no haber excedentes alimentarios), sin especialización económica ni productiva.

A medio camino entre lo biológico y lo cultural, la normatividad prejurídica, de naturaleza moral, opera en sociedades en las que se concibe a cada uno de los miembros con una individualidad atenuada, relativa,

[...] como integrante de un grupo unido por una solidaridad extremadamente estrecha en que se piensa en términos de unidad colectiva [...] en la que los mitos, los rituales y la educación contribuyen a integrar al individuo en una colectividad casi orgánica con una consciencia colectiva (Bernal, 2011: 174).

No es, pues, posible localizar en estas sociedades prejurídicas el concepto de persona en el sentido jurídico objeto de nuestro estudio: como sujeto de derechos. En ellas, el individuo aparece como un ser moral copartícipe de una consciencia colectiva y, al mismo tiempo, como un ser autoconsciente que reconoce a sus congéneres como iguales a los que reprochar el incumplimiento de la norma grupal. No obstante, no aparece todavía como titular de un conjunto de derechos y obligaciones exigibles frente a otro o frente a una autoridad reguladora.

1.1.3. Sociedades protojurídicas

El paso de aquella normativa prejurídica a los sistemas protojurídicos puede vincularse a la aparición de excedentes alimenticios susceptibles de almacenamiento, que trajo consigo un aumento de la densidad poblacional, así como la especialización y la división del trabajo (Bernal, 2011: 199).

En grupos de millares de personas sin vínculos de parentesco, los intercambios recíprocos son insuficientes para atender a las necesidades del grupo, y la gestión de excedentes exige su redistribución. Esta función acaba recayendo en una jefatura centralizada de carácter hereditario. Esta jefatura asume, junto con el poder de redistribuir los excedentes, la potestad del uso de la fuerza y la facultad de dictar normas a la colectividad o canalizar las reveladas por la divinidad o las utilizadas por los antepasados.

Paralelamente, la moral, el *ethos* colectivo, deja de ser (por ineficaz) un medio de resolución de conflictos. Este papel corresponde ahora a un conjunto normativo, consuetudinario, no escrito, regulador de derechos y obligaciones para los miembros de la comunidad y, como producto cultural, en esencia mutable. Además, en estas estructuras sociales más complejas encontramos el tránsito del fenómeno natural de la apropiación de las cosas con ánimo de pertenencia al verdadero derecho de propiedad y, con ello, el reconocimiento del individuo como su titular.

1.1.4. Sociedades jurídicas plenas

Las sociedades jurídicas plenas son aquellas que reconocen tanto las normas jurídicas reguladoras de relaciones entre sus miembros como las normas reguladoras del procedimiento para el establecimiento y la aplicación de aquellas. Su consolidación es paralela a la aparición de la escritura y a una estructura social compleja, estratificada, con una división formal y especializada del trabajo y un poder centralizado con monopolio para el uso de la fuerza, la promulgación de leyes y la resolución de conflictos. Estas primeras organizaciones sociales complejas (calificadas como estados prístinos) se asientan en Mesopotamia (3300 a. C.), Egipto (3100 a. C.), valle del Indo (2000 a. C.), cuenca del río Amarillo (2000 a. C.), este de África (200 d. C.) y Mesoamérica (300 d. C.) (Harris, 1986: 83-104).

La consolidación del Derecho como mecanismo regulador supone que las normas surjan ahora de un acto humano voluntario, individual o colectivo (dependiendo de la forma de gobierno), y no de la reiteración de conductas éticas por la colectividad. A partir de ahora, *Homo sapiens* vivirá para siempre

integrado en una red de contratos y normas; en resumen, de relaciones jurídicas, constituyéndose en *Homo juridicus* (Supiot: 2007).

1.2. LOS DESTINATARIOS DE LAS PRIMERAS NORMAS JURÍDICAS

1.2.1. Asia Menor, Egipto y Grecia

El surgimiento del Derecho como mecanismo regulador de las relaciones humanas no provocó la aparición simultánea de un concepto jurídico de persona y de una categoría unitaria de sujeto de derecho; todo ello fue fruto de la dogmática jurídica posterior, especialmente a partir del Derecho romano. No obstante, es posible extraer de las normas jurídicas primitivas una serie de características de sus destinatarios, que se detallan a continuación.

a) Los obstáculos terminológicos para su determinación

Una de las dificultades comunes al estudio de los cuerpos legales más antiguos (códigos de Ur-Nammu, Eshunna y Hammurabi, los Decretos de Palacio asirios y los códigos hititas) deriva de la traducción de los términos referidos a los sujetos de sus disposiciones (Roth, 1995: 79). Desde luego, las referencias al término «persona» en los textos mesopotámicos solo pueden entenderse en el sentido genérico de hombre como sujeto indeterminado, equivalente a cualquiera, a alguien o a alguno, y no como alusiones a la idea de sujeto de derecho o, menos aún, ciudadano.

Igualmente, en el antiguo Egipto son varias las palabras que, según los períodos y contextos, sirven para definir al hombre libre, pero no es seguro que ninguno de estos términos considere al individuo así designado como un ciudadano, aunque se regula ya la participación de hombres y mujeres en la administración municipal con derecho a voto en los tribunales locales (Husson y Valbelle, 1998: 136).

Será en la antigua Grecia donde se consagre el concepto de ciudadanía como estatuto que habilita para la participación en la vida pública, pero hay que advertir que tampoco los textos griegos (ni filosóficos ni legales) reconocieron una figura similar a nuestros derechos. Esto se debe a que Grecia nunca situó al individuo como epicentro del Derecho, sino que más bien los bienes e intereses individuales se integraban en el interés común y eran garantizados a través de la protección de este (Megías, 2016: 65). Puede, por lo tanto, afirmarse que la aportación griega al concepto de persona (*πρόσωπον*, *prosopon*) es, ante todo, etimológica.

b) La estratificación y jerarquización social

A falta de aquellas categorías conceptuales, los primeros cuerpos legales establecen una regulación casuística y asistemática de diversas cuestiones civiles, penales y administrativas, en las que la categoría social del sujeto es la medida de la pena, indemnización o castigo.

Mesopotámicos y asirios distinguirán entre personas libres, siervos y esclavos, y el género, la edad, el orden de nacimiento, e incluso la profesión, determinarán también una fuerte jerarquización de las relaciones jurídicas, así como consecuencias jurídicas diversas ante hechos idénticos.

Características similares encontramos en las antiguas leyes egipcias, en las que, además, se perfila el estatuto jurídico de los extranjeros, llegados a Egipto o bien como inmigrantes, contratados como soldados o siervos, o bien como prisioneros de guerra. Respecto a los esclavos, en atención al grado de sometimiento al señor, encontramos diversas categorías, algunas equiparables a la de los siervos de la gleba de la Edad Media —en relación de feudalismo, vinculados a la tierra—, otras asimiladas a empleados domésticos y otras idénticas a la esclavitud romana: son objeto de compra-venta y alquiler, y, aunque se ignoran las formalidades jurídicas, debió de existir en el antiguo Egipto un acto similar a la manumisión romana, pues hay constancia de numerosos esclavos liberados e integrados en la vida civil con el estatus de libres.

c) La imputación de consecuencias jurídicas a entidades distintas del ser humano

Por último, y aunque pueda ser cierta la afirmación de que para contestar cumplidamente la pregunta de cómo se concebía la persona jurídica en la Antigüedad se necesitaría dedicarle toda una vida, es destacable que ya en los primeros textos legales (códigos de Hammurabi y Eshunna) encontramos también referencias a entes personificados (dioses, palacios) como titulares de derechos de propiedad y de crédito por indemnizaciones por daños o robos.

De forma más concreta y directa, las leyes egipcias reconocieron aptitud legal para la titularidad de derechos a las fundaciones funerarias, desmintiéndose así la opinión doctrinal mayoritaria que suele situar en el Derecho romano el origen de la fundación como persona jurídica (Martin y Lara, 1999: 60).

1.2.2. Roma

Al igual que ocurrió en las civilizaciones anteriores, en el Derecho romano no todo ser humano es sujeto de derechos, sino solo el hombre libre, ciudadano romano y *sui iuris*, circunstancias (libertad, ciudadanía y no sometimiento a autoridad familiar) que solo concurrían en el *pater familias*. De esta forma (Iglesias, 1983: 118) la personalidad, es decir,

[...] la capacidad jurídica del hombre —del individuo en sí, con independencia de cualquier otra condición—, así como de las organizaciones humanas —personas jurídicas—, solo fue reconocida en época avanzada del Derecho Romano. [...] En el decurso de los tiempos permaneció firme la idea de vincular la capacidad al hombre libre. No ocurrió así con las otras condiciones —la de ser *civis romanus* y *sui iuris*— que fueron paulatinamente desviadas.

Los juristas clásicos romanos no atribuyeron un sentido técnico al término «persona», si bien en sus escritos hay frases que se han interpretado como indicios para perfilar aquel sentido técnico, pues junto a textos en los que se utiliza como sinónimo de «hombre», en otros se menciona para negar esa cualidad de persona al esclavo (*persona non est* D. 36, 1, 57), y en algunos se emplea como equivalente a la cualidad con que actúa el hombre, a su posición ante ciertas situaciones o para referirse a las ciudades, al pueblo romano o las colonias. También encontramos en Roma la atribución de subjetividad jurídica a algunos dioses que, como Apolo o Júpiter, podían ser instituidos herederos (Ulp. 22, 6).

Tampoco consagró Roma en la época clásica un nombre técnico para designar la personalidad o la capacidad jurídica en cuanto aptitud general para ser sujeto titular de derechos y obligaciones, pues el vocablo *capax* se vincula a la aptitud singular para entrar en el juego de relaciones determinadas y concretas, faltando aquella noción de generalidad (Iglesias, 1983: 117). Asimismo, la palabra *caput* significa ‘cabeza’, ‘individuo’, pero se vincula tanto al hombre libre como al esclavo, abstracción hecha de su capacidad.

Puede, pues, concluirse que, en el Derecho romano, *persona* denota de manera abstracta la subjetividad jurídica y no el sujeto, por breve que haya sido el paso de cambiar la cualidad con el ente cualificado (Ferrara, 2006: 236).

1.3. LA CONCRECIÓN DEL CONCEPTO

Hemos visto cómo el concepto jurídico de persona nace históricamente como una cualidad jurídica vinculada no a la naturaleza humana, sino a la idea de estatus, hecho que se reproduce también en los Derechos germánicos, que, sin forjar un concepto doctrinal sobre la persona, reconocieron aptitud para ser sujeto de derechos tan solo al hombre libre, perteneciente al pueblo, dejando así fuera a los esclavos, a los extranjeros y a los condenados por ciertos delitos («pérdida de la paz» era la expresión que utilizaban los pueblos germánicos para designar esa muerte civil).

Esta idea, la distinción entre la condición de ser humano, procedente de la naturaleza, y la cualidad de persona como atributo jurídico procedente de la ley, aún subsiste en los siglos XVI y XVII: «*Servus cum homo est, non persona; homo naturae persona iuris civilis vocabulum*», escribió Hugues Doneau (Doneau 1846: 241).

En el paso a la equiparación técnica (no meramente vulgar) de ambos conceptos —persona y ser humano— será decisiva la aportación de la patrística y de la escolástica cristianas, que en el curso de sus discusiones sobre la naturaleza de la Trinidad, en especial a partir del Concilio de Nicea, la concibió como una esencia (*ousía*) con tres personas o entes (*hipóstasis*). Es en este marco en el que Boecio (calificado como el último romano y el primer escolástico), para destacar la naturaleza dual, humana y divina de Cristo, concibe a la persona como sustancia individual de naturaleza racional. Esta idea será elevada a definitiva por santo Tomás (*Summa Theologica* I, 29), al señalar que con esa denominación, *persona*, se alude a la dignidad máxima existente en la naturaleza y que deriva de la racionalidad.

La última aportación al concepto de persona vendrá de la mano del humanismo filosófico, que vincula aquel elemento de racionalidad a la idea del libre albedrío humano. Rousseau, en el capítulo II del libro I de *El contrato social o principios de derecho político*, afirma que llegado el hombre al estado de la razón y «siendo el hombre juez de los medios que lo han de conservar, se convierte por este hecho en dueño de sí mismo». Trasladando estas ideas al universo jurídico, Leibniz vincula los conceptos de persona y voluntad, Hegel los relaciona con la idea de autoconciencia y, ya a principios del siglo XX, Wundt refiere el concepto a la unidad de sentir, pensar y querer.

Sobre estos cimientos (libertad, libre albedrío y racionalidad) descansa el concepto de persona que finalmente cristaliza en la *Declaración Universal de los Derechos Humanos* del 10 de diciembre de 1948, en cuyo preámbulo se proclama que la libertad, la justicia y la paz en el mundo tienen como base el reco-

nocimiento de la dignidad intrínseca y de los derechos iguales e inalienables de todos los miembros de la familia humana.

Este recorrido histórico por el concepto de persona pone de relieve que se ha construido en torno a tres elementos: el antropológico, el filosófico y el jurídico. Sus posibles combinaciones nos ayudan a comprender las distintas posiciones doctrinales existentes al respecto: el elemento antropológico nos sitúa ante el ser humano; el filosófico, ante un ente racional, consciente o sintiente (según las distintas corrientes); y el jurídico, ante un ente con función jurídica en un sistema normativo.

De esta forma, las posiciones llamadas realistas ubican el centro de gravedad del concepto en el aspecto antropológico y vinculan el concepto de persona con el ser humano de manera exclusiva y excluyente. Esta es la tesis seguida, entre otros, por Hölder (citado por Ferrara: 2006, 238), quien define a la persona como la «individualidad específica humana», así como por De Castro (2008: 30), para quien el concepto de persona descansa en «el reconocimiento de la situación jurídica que corresponde (*dignitas*) a todo hombre por su condición de ser racional creado a imagen y semejanza del Creador». Desde esta perspectiva se ha sostenido que el concepto de persona es un *prius* para el Derecho, que se limita a reconocer esa cualidad previa, natural e inherente a todo ser humano. Así, Castiella (2003: 19), citando a Spaemann, afirma que

[...] no existe un tránsito paulatino desde «algo» a «alguien». De algo no deviene alguien. Si el ser persona fuera un estado, podría surgir poco a poco. Pero si persona es alguien que pasa por diferentes estados, entonces, los supone todos [...]. No comienza a existir después del hombre ni se extingue antes que él [...] el ser personal no es el resultado de un desarrollo sino la estructura característica de un desarrollo que mantiene una unidad a través del tiempo. Esa unidad es la persona.

Esta argumentación lleva a algunos sectores doctrinales a reclamar el reconocimiento del embrión humano como persona y a la consideración de los entes jurídicos no humanos no como verdaderas personas, sino solo en un sentido traslaticio, de modo que las llamadas personas jurídicas y los sistemas de inteligencia artificial solo tendrían esta consideración por vía analógica o referencial. Se trataría de meras personificaciones por vía de ficción, y nunca de verdaderas personas.

Desde una óptica filosófica, situar el eje del concepto de persona en su naturaleza racional y concebirla como un sujeto autoconsciente al que se le exige, según explica Troiano (2006: 238, citado por Ferrara), «plenitud de conciencia de su propia individualidad» llevaría a cerrar aquel concepto a los ni-

ños, que serían solo personas en potencia, y a los discapacitados cognitivos, al tiempo que lo abriría a los (hoy todavía hipotéticos) sistemas algorítmicos autoconscientes. Peter Singer (Polo 2004: 65), desde una perspectiva utilitarista, deslinda de forma clara los conceptos de persona y ser humano, reservando la primera categoría para el ser racional y consciente capaz de concebirse a sí mismo como una entidad distinta existente en el tiempo, y dejando la calificación de ser humano como un criterio puramente biológico de pertenencia a una especie. En lo que se refiere (Singer 1995: 107-109) a esta última cuestión

[...] determinar si un ser es miembro de una especie concreta es algo que se puede hacer científicamente examinando la naturaleza de los cromosomas en las células de los organismos vivos. En este sentido no existe duda de que desde los primeros momentos de su existencia un embrión concebido de un óvulo y un espermatozoide humano es un ser humano y lo mismo ocurre con el ser humano que se encuentre discapacitado psíquicamente de la manera más profunda e irreparable o incluso un bebé que haya nacido sin cerebro.

Pero en lo que respecta al concepto de persona, enlaza, según el propio Singer, con lo que el teólogo protestante Joseph Fletcher (citado por aquel) «denomina indicadores de la condición humana entre los cuales se encuentran los siguientes: conocimiento y control de uno mismo, sentido del futuro, sentido del pasado, capacidad de relacionarse con otros, preocupación por los demás, comunicación y curiosidad».

Frente a las concepciones anteriores, los defensores de las teorías formalistas ven en la persona una categoría jurídica abstracta desvinculada de toda idea de corporalidad y de elementos volitivos e intelectivos. Lo esencial, desde esta óptica, es la cobertura formal que atribuye el ordenamiento jurídico para actuar en él, con independencia de que el sustrato al que se le atribuya sea el ADN de la especie humana, la autoconsciencia o el mero reconocimiento legal. En esta línea, explica Ferrara (2006) que no trocar los diferentes significados del término «persona» ni utilizar con fines jurídicos conceptos y principios ajenos al Derecho es una exigencia metodológica. La biología determinará con precisión la especie a la que pertenece un determinado individuo, pero la persona es una categoría jurídica que por sí misma no implica ninguna condición de corporalidad, sino que ser punto de confluencia de consecuencias jurídicas basta formalmente para que haya un sujeto de derechos, concibiéndose así la persona como un centro de imputación de normas jurídicas.

Los excesos de las teorías formalistas, derivados sobre todo del abuso de la persona jurídica (la utilización de su cobertura con ánimo defraudatorio) se

tradujeron en una crisis del concepto de persona jurídica como una construcción dogmática unitaria. Asimismo, se ha destacado que las construcciones doctrinales antes expuestas «consideran solo un aspecto de una misma realidad» (Fernández, 2001: 18), es decir, las teorías realistas y utilitaristas incidirían solo en el aspecto ético-jurídico del problema, olvidando sus aspectos dogmáticos, y la teoría formalista se preocuparía tan solo del ángulo normativo, dejando de lado la realidad existencial. Y es que, aunque la mera cobertura formal atribuida por el ordenamiento jurídico pueda ser suficiente para la creación de personas jurídicas, es necesario (lo impone la lógica interna del sistema) que esa cobertura sea compatible con los fines del propio ordenamiento jurídico y, dado que este surge como un producto cultural regulador de conductas y relaciones entre seres humanos, solo lo será en la medida que esa cobertura formal sirva a la realización de intereses humanos. Con ello se pone de relieve que el ser humano, todo ser humano, ha de ser persona para el Derecho, pues este es una institución humana y el hombre es el átomo de la vida jurídica, de modo que, desde una óptica kantiana, es sujeto de derecho en cuanto es un fin en sí mismo. Al mismo tiempo, no cabe decir que las personas jurídicas son una mera invención de la ley, sino la forma que la ley atribuye a ciertos fenómenos (corporativos, asociativos, fundacionales) que existen en la realidad social y a través de los cuales el ser humano desarrolla sus fines.

Esquemmatizando ahora las combinaciones de aquellos tres elementos —antropológico, filosófico y jurídico— en torno a los cuales se ha construido el concepto de persona, podemos distinguir los siguientes modelos (en los que se señala con (i) la presencia del elemento y con (o) su ausencia):

- a) Persona = ADN humano (i), ente racional/consciente/sintiente (o) y ente con función jurídica (i): a este esquema responde la práctica totalidad de los ordenamientos jurídicos modernos, en los que la condición de persona se le atribuye al ser humano una vez cumplidos los requisitos legalmente establecidos («nacimiento con vida, una vez producido el entero desprendimiento del seno materno» en el caso del Derecho español, artículo 30 del Código Civil) y a las personas morales o jurídicas que se constituyen cumpliendo los requisitos formales que establece cada ordenamiento jurídico.
- b) Persona = ADN humano (i), ente racional/consciente/sintiente (i) y ente con función jurídica (o): anticipando por la vía especulativa en el ámbito social lo que Darwin demostraría seis años después para lo biológico, Herbert Spencer acuñó la frase «la supervivencia del más apto», abriendo así el paso a las corrientes gradualistas de la concepción de la persona, que exigen

la concurrencia en el ser humano de plenitud de facultades mentales para su consideración como persona.

- c) Persona = ADN humano (1), ente racional/consciente/sintiente (1) y ente con función jurídica (1): un ejemplo normativo de la concurrencia de estos tres elementos (superponiendo legalmente el pleno raciocinio a la condición humana para ser considerado persona jurídicamente) lo encontramos en el llamado Decreto de Eutanasia de Adolf Hitler, en vigor desde el 1 de septiembre de 1939 y a cuyo amparo fueron asesinados de manera legal al menos 70 000 discapacitados mentales cuyas vidas «no eran dignas de ser vividas».
- d) Persona = ADN humano (0), ente racional/consciente/sintiente (0) y ente con función jurídica (1): en este modelo podrían situarse las posiciones jurídico-formales que construyen el concepto de persona única y exclusivamente en torno al reconocimiento formal que atribuye el ordenamiento jurídico.
- e) Persona = ADN humano (0), ente racional/consciente/sintiente (1) y ente con función jurídica (0): diversas concepciones filosóficas han situado como punto esencial de la persona su naturaleza racional, o su facultad de autoconciencia o su naturaleza sintiente; más tarde, al tratar de la comunidad moral, se hará referencia a alguna de ellas, entre las que se incluyen las pretensiones animalistas de concesión de derechos a los animales.
- f) Persona = ADN humano (0), ente racional/consciente/sintiente (1) y ente con función jurídica (1): no existen ejemplos normativos de sistemas jurídicos que atribuyan la condición de persona en atención a las capacidades de raciocinio, autoconciencia o naturaleza sintiente y haciendo abstracción de la pertenencia a la especie humana, aunque podrían ubicarse en este esquema las pretensiones incluidas en la Declaración de los Grandes Simios si jurídicamente la categoría de homínido llegase a tener sustantividad (como postula el Proyecto Gran Simio).
- g) Persona = ADN humano (1), ente racional/consciente/sintiente (0) y ente con función jurídica (0): a este esquema responderían las posiciones que consideran, con independencia del reconocimiento del ordenamiento jurídico, que el ser humano es persona desde el mismo momento de la concepción.
- h) Persona = ADN humano (0), racional/consciente/sintiente (0) y ente con función jurídica (0): la consideración como persona de un ente no humano, irracional y sin reconocimiento jurídico actual encaja en la concepción dataísta a la que se hará referencia en un momento posterior de este trabajo. Se trata de la personificación de un algoritmo artificial con relevancia jurídica para el ordenamiento.

Se le suele atribuir a Leibniz la creación de la expresión «sujeto de derecho», combinando la concepción de Grocio del derecho subjetivo como una cualidad moral y la idea de la naturaleza racional humana. Por lo general, en un sentido técnico-jurídico, los términos «persona» y «sujeto de derechos» se emplean como sinónimos.

Jurídicamente, «sujeto de derecho» se entiende como el vocablo «derecho», en un doble sentido: «sujeto del derecho» para hacer referencia al elemento subjetivo del ordenamiento jurídico (el derecho objetivo, las normas) y «sujeto de derechos» para referirse a la posición de la persona como titular de derechos subjetivos, es decir, como centro subjetivo de imputación de derechos y obligaciones derivados de la relación jurídica (el derecho como potestad).

Ocurre entonces que la concepción que tengamos del término «derecho subjetivo» predeterminará también el concepto mismo de persona y, así, según las distintas tesis resumidas por Ferrara (2006: 242, 247), se ha sostenido:

- 1) Que el derecho subjetivo es un poder de la voluntad, una esfera de autonomía que el ordenamiento jurídico pone a disposición del sujeto. Esta teoría clásica, defendida por Savigny y Windscheid, supondría negar a los niños o discapacitados cognitivos aptitud para ser titulares de derechos, y queda desvirtuada al observar que el ordenamiento jurídico tutela derechos en ausencia de la voluntad del titular, e incluso, en ocasiones, en contra de esta.
- 2) Que el derecho subjetivo es un interés jurídicamente protegido, tesis también clásica defendida por Ihering y que, llevada al extremo, implicaría calificar como persona a todo lo que tenga un interés relevante para el Derecho: animales, plantas u objetos valiosos. Esto equivale, explica gráficamente Francesco Ferrara, a confundir lo protegido (el interés) con los mecanismos de protección (de los que el derecho subjetivo es una de las modalidades pero no la única).
- 3) Que el derecho subjetivo es el interés protegido por el reconocimiento jurídico de la voluntad, fórmula mixta que, a base de combinar las dos tesis anteriores, es criticable por las mismas razones que aquellas.

- 4) Que el derecho subjetivo es sencillamente un efecto del Derecho objetivo individualizado. Según esta tesis, lo que ocurre es que el ordenamiento jurídico impone, a través de la aplicación de sus normas, deberes jurídicos en interés de los particulares, concediéndoles la potestad de obtener su defensa en caso de vulneración de aquellos. De ahí se desprende, concluye Ferrara, que «persona, sujeto de derechos» significa jurídicamente ‘titular de deberes y poderes jurídicos’.

Expuesto lo anterior, puede procederse al análisis de quién es persona, sujeto de derecho en los ordenamientos jurídicos modernos. Desde luego, en este punto, sigue vigente el *Digesto* cuando proclamó que «hominum causa omne ius constitutum est» (D. 1, 5, 2), y dado que el Derecho nació y existe para y por los hombres, es evidente que el ser humano es la respuesta inmediata a aquella cuestión.

2.1. RAZONES

2.1.1. La comunidad moral

La historia del Derecho puede concebirse como la historia de la ampliación de una comunidad moral, ya que, en estadios primitivos, «los sentimientos de respeto, simpatía y solidaridad y, consiguientemente, la consideración moral, se limitan al grupo inmediato al que uno pertenece» (Mosterín, 2015: 56), pero el círculo moral de respeto y solidaridad se ha ido extendiendo para abarcar grupos cada vez más amplios. No puede negarse «que en ocasiones ha habido dificultades cuando se trataba de dar el salto a otra religión, a otra raza o a otra nación» (Mosterín, 2015: 56), pero las declaraciones de derechos derivadas del proceso ilustrado y las revoluciones del siglo XVIII, más tarde acogidas y reformuladas por la comunidad internacional, cristalizaron en el concepto de derechos humanos. De este modo, en la época moderna, la consideración del ser humano, sin distinción alguna, como sujeto de derechos es un presupuesto esencial de todo ordenamiento jurídico y no precisa un reconocimiento expreso.

2.1.2. Contractualismo social

El concepto de derechos humanos surge en el marco del iusnaturalismo, siendo esencial la aportación de John Locke, que sitúa la ley natural, en cuanto ley

racional del estado de naturaleza, como fundamento de los que califica como derechos naturales del hombre (considerando, como tales, la vida, la libertad y la propiedad), y estos como origen de la sociedad y de la comunidad política. Y así, defendiendo la igualdad como atributo natural del hombre, afirma (Locke, 2005: 19) que

[...] el estado de naturaleza tiene una ley que lo gobierna y que obliga a todos; y la razón, que es esa ley, enseña a toda la humanidad que quiera consultarla que siendo todos los hombres iguales e independientes, ninguno debe dañar a otro en lo que atañe a su vida, salud, libertad o posesiones.

Y, sobre esa base, Locke articula la comunidad política como una renuncia de cada uno de los individuos (en favor de la comunidad) al poder que les otorga la ley natural.

Esta tesis, la existencia de un estado de naturaleza o prejurídico que se supera en virtud de un acuerdo que determina los principios de la organización jurídica y política, es común a las teorías contractualistas de legitimación de la comunidad política, según las cuales dicho acuerdo vendría a poner fin a una situación de «guerra de todos contra todos», en palabras de Thomas Hobbes. Es evidente que la hipótesis del contrato social tiene un encaje imposible en el proceso histórico bio-cultural de hominización y socialización (expuesto al principio de este trabajo), más allá de servir de analogía vaga y puramente referencial, y ya Kant y Gierke destacaron que el contrato social no se dirige a explicar el origen histórico del orden socio-político, sino su fundamento, pues no se trata de un hecho (Kant, 1999: 31), sino «de una mera idea de la razón, que tiene, sin embargo, su indudable realidad (práctica) en cuanto obliga al legislador a que dé sus leyes como si estas *pudieran* haber emanado de la voluntad del pueblo».

La idea central del contractualismo social radica en establecer un procedimiento a través del cual, mediante la negociación, se producen renunciaciones a los poderes individuales en favor del poder de la ley y de la organización política. Cuestiones esenciales para la óptica contractualista son la determinación de la causa (el fundamento del contrato) y de los sujetos negociadores del contrato.

La necesidad de tutela de los intereses individuales es una tesis tradicional a la hora de fundamentar el contrato social, si bien no explica qué incentivos obtendrían para celebrar el contrato los sujetos que, por su fortaleza, pueden imponer siempre sus intereses en el estado de naturaleza, aunque, ciertamente, se ha sostenido (Hart 1994: 194 y ss.) que la vulnerabilidad afecta a todos los seres humanos, y eso sería un incentivo suficiente para acordar la prohibición

de la violencia. Esta tesis tradicional tampoco explica de manera adecuada por qué los sujetos (como menores y discapacitados) carentes de capacidad negociadora en el contrato social son incluidos en la comunidad resultante si el contrato se dirige a la tutela de los intereses propios de los sujetos que lo suscriben y acaba concibiendo a aquellos sujetos como excepciones o casos marginales.

En la época moderna, el contrato social se ha justificado sustentándose en la idea de imparcialidad que inspira a quienes lo suscriben. John Rawls, en su *Teoría de la Justicia*, es uno de los más firmes defensores de esta tesis. Rawls propone un experimento mental a partir de una posición original (remedo del clásico «estado de la naturaleza») en la que los seres humanos estipulan por contrato los principios de justicia en que se estructurará la vida social. Para evitar las posiciones de ventaja que (por fuerza o talento) ciertos individuos pueden tener en esa situación, el acuerdo exige, por un lado, que todos los participantes sean *rough equals* (iguales en sentido bruto, con un mínimo idéntico de capacidades mentales y físicas) y, por otro, la interposición de un «velo de ignorancia» que oculta todos los hechos (propios y ajenos) que puedan predeterminar el concepto de justicia, para así garantizar la imparcialidad en el acuerdo. Ahora bien, aunque el velo de ignorancia implica que los individuos que negocian en el estado original (Cuenca 2012: 108) ignoren cuál será su género, su raza, o su clase social, la participación en esa negociación exige (Rawls 1995: 31) tener capacidad para conocer el sentido de la justicia y para concebir lo que es bueno para uno mismo, por lo que los seres humanos que no reúnan esos requisitos (como los menores y los discapacitados) no son *rough equals* y, por lo tanto, no son sujetos de justicia que puedan participar en el contrato. Así, en la construcción de Rawls (Cuenca 2012: 108), el trato moral a menores y discapacitados,

[...] es una «cuestión práctica urgente» que deberá ser tratada en un momento posterior, en el estadio legislativo, pero no una cuestión básica de justicia que debe ser abordada en un primer momento. Las personas con discapacidad son tratadas como «ciudadanos de segunda», tanto por lo que respecta a su consentimiento, como por lo que respecta a sus demandas e intereses.

De esta forma, se pone de relieve que el contrato social no justifica por sí solo la comunidad moral, pues dejaría sin cobertura a los llamados «casos marginales» (discapacitados mentales, enfermos en coma, niños de corta edad). Es, por lo tanto, necesario acudir a algún elemento adicional más allá del contrato social, de forma que tales individuos están incluidos en la comunidad moral

por la mera pertenencia a la especie humana, que, según se ha explicado desde ópticas iusnaturalistas, inviste a todo ser humano de una especial dignidad y una serie de derechos inherente a ella, de manera antecedente y por naturaleza.

2.1.3. Utilitarismo

Las corrientes utilitaristas acuden a vías distintas a las del contrato social para justificar las comunidades moral y jurídica, partiendo de un doble planteamiento común: el consecuencialismo como criterio metodológico y el principio de utilidad como criterio valorativo.

La moral de principios (reglas) propia de la tradición iusnaturalista se enfrenta ahora a una moral de resultados que rompe de un modo radical con la idea de una humanidad investida de manera natural de derechos. Explica Mosterín (2015: 59) que Jeremy Bentham calificó la ley natural de «sinsentido» y los derechos naturales (pretendidamente derivados de aquella) como «sinsentido sobre zancos». Su idea del principio de utilidad gravita en torno a la maximización del placer y la minimización del dolor, principio coherente con la pregunta que Bentham se formula para determinar quiénes pueden ser sujetos de derechos: «la cuestión [ya] no es ¿pueden razonar?, ni ¿pueden hablar?, sino ¿pueden sufrir?».

Esta idea abre la puerta a la distinción entre persona y ser humano, y será llevada en los tiempos modernos a sus últimas consecuencias por Peter Singer, quien, a partir del denominado utilitarismo de preferencias, señalará que la pertenencia a la especie *Homo sapiens* no es un criterio moral, no tiene «pertinencia moral», sino tan solo biológica.

El utilitarismo de preferencias sostiene que toda acción contraria a las preferencias de cualquier ser es mala, moralmente reprochable, a no ser que existan preferencias contrarias que tengan más peso que esta (Polo, 2004: 61-78), y el criterio para la posesión de preferencia lo da la sintiencia (aquella capacidad de sentir, de experimentar placer y dolor), de ahí que, para esta orientación, ese criterio sea el determinante para la integración en la comunidad moral.

De nuevo, se observa que las tesis utilitaristas quiebran al tratar los llamados casos marginales. En efecto, así como las teorías de corte contractualista adolecen de falta de coherencia interna en cuanto deben acudir a un argumento distinto de la convención social para incluir en la comunidad moral a quienes por falta de capacidad no podrían formar parte de ella, las tesis utilitaristas chocan con la realidad normativa de todos los ordenamientos jurídicos modernos, que atribuyen a niños y discapacitados mentales la condición de sujeto de

derechos y recurren también a otros argumentos para justificar la inclusión como sujetos de derechos a quienes carecen de autoconciencia (entendida esta en el sentido de capacidad de autorrepresentación en el tiempo).

Así, Hugo Tristram Engelhardt (1995: 147-149) admite el choque de estas teorías utilitaristas con la realidad social cuando argumenta que el (universal) reconocimiento social y jurídico de menores y discapacitados mentales como sujetos de derecho no puede ser justificado en función de la gramática básica de la moralidad, debido a que dichas entidades no tienen una categoría intrínsecamente moral por no ser agentes morales. Por ello, «tendremos que justificar un sentido social de la persona, más bien en función de la utilidad que tiene la práctica de tratar a ciertas entidades como si fueran personas». Engelhardt (1995) explica que el rol social de persona atribuido a embriones, bebés y discapacitados desprovistos de autoconciencia puede justificarse porque representa valores o virtudes esenciales para la vida social, como pueden ser la compasión, la solidaridad y la protección de la vida humana, en especial cuando esa vida humana es frágil e indefensa. Además, con respecto a niños y a otros humanos *ex utero*, la atribución del estatus de persona es una salvaguarda contra la incertidumbre de cuándo exactamente los seres humanos se convierten en personas en sentido estricto. Al mismo tiempo, asegura la crianza de los niños, mediante la cual los seres humanos se convierten en personas en sentido estricto.

Ello supone reconocer, por un lado, un concepto de persona en sentido estricto —el que define al sujeto como agente moral— y, por otro, hasta tres sentidos de persona en sentido social, que permiten otorgar derechos y tratar como pacientes morales a seres que no son propiamente personas (como los niños), a individuos que alguna vez fueron personas pero que ya no lo son (como los enfermos de Alzheimer) y a individuos que nunca han sido ni nunca serán personas en sentido estricto (como los discapacitados mentales).

En rigor, el principal escollo de las tesis utilitaristas de la persona es su individualismo (Hall, 2015: 77), pues estas teorías aplican el criterio de preferencias a cada individuo, y por eso se encuentran con el problema de cómo fundamentar la atribución de derechos morales a sujetos carentes de preferencias (por ejemplo, recién nacidos y discapacitados mentales). Este problema se evita si se toman en consideración no las aptitudes concretas de cada individuo, sino las capacidades generales que perfilan el funcionamiento normal de la especie, aplicando criterios generales a todos los seres humanos e incluyendo, entre ellos, los casos excepcionales que no cumplan aquellos criterios, pues, al fin y al cabo, las leyes no regulan casos individuales, sino que se caracterizan por su generalidad. Tal es el planteamiento propuesto por Martha Nussbaum

en su obra *Las fronteras de la justicia*. Ocurre, sin embargo (Cuenca, 2012: 109-111), que aunque dicha teoría parte de la consideración de los discapacitados como miembros plenamente iguales de la comunidad humana, al delimitar las capacidades básicas «que dan forma y contenido a la idea abstracta de dignidad» humana

[...] el enfoque de Nussbaum se muestra mucho menos inclusivo de lo que a priori se declara, al menos para algunas personas con discapacidad. Su esquema falla a la hora de reconocer la plena dignidad de las personas con discapacidad cuyo nivel de funcionamiento en las capacidades básicas se sitúa por debajo del umbral mínimo universal y excluye a estos individuos de la participación plena en la sociedad.

Vistas estas posiciones teóricas sobre el fundamento de la comunidad moral, ahora procede hacer referencia a los supuestos bajo los cuales, en la práctica, los ordenamientos jurídicos han modificado su contenido por defecto o por exceso.

2.2. DEFECTOS

2.2.1. Esclavos, mujeres y enemigos

Hasta épocas muy recientes, la condición de sujeto de derechos estaba vetada formalmente y con respaldo legal a cierta categoría de seres humanos, desmintiendo así por la vía fáctica el ideal iusnaturalista de equiparación exclusiva y excluyente entre ser humano y persona. Así, por ejemplo, en 1857, el Tribunal Supremo de Estados Unidos afirmaba en la sentencia *Dred Scott vs. Sandford* que «los negros son seres de una categoría inferior que no tienen ningún derecho que el hombre blanco deba respetar», reconociendo así la competencia de los estados para legislar sobre la esclavitud. La misma sociedad estadounidense que proclamó, en el preámbulo de su Constitución, la finalidad de «establecer Justicia [...] promover el bienestar general y asegurar [...] los beneficios de la Libertad» no tuvo ningún inconveniente en asentar, en el artículo 4 de la misma norma suprema, la esclavitud como uno de los cimientos de esa justicia, libertad y bienestar, al tiempo que computaba a los seres humanos de raza negra como «tres quintos de persona».

De forma similar, el Code civil francés de 1804, al tiempo que se presentaba al mundo como el crisol de los principios revolucionarios que garantizaban

la libertad individual, poniendo fin de forma expresa al régimen feudal, proclamaba sin sonrojo, al definir las cosas semovientes en su artículo 528, que, además de los animales que cita por vía de ejemplo, existían otras cosas que podían desplazarse por sí mismas, en referencia a los esclavos.

En realidad, la idea romana de que el hombre es la razón de la existencia del Derecho nació quebrada. El Derecho romano mismo, al igual que todos los ordenamientos jurídicos anteriores, reguló la institución de la esclavitud, perfilando el estatus jurídico de un ser humano desprovisto de la cualidad jurídica de persona; privado, pues, de la aptitud para ser titular de derechos y obligaciones y, al contrario, catalogado no de sujeto, sino de objeto de derechos. «*Servile caput nullum ius habet*» (*Digesto*, 4, 5, 3, 1).

De igual modo, la condición jurídica de la mujer, en muchas épocas y lugares, ha quedado excluida de la condición de persona para el Derecho, o sus derechos o la posibilidad de su ejercicio, bastante reducidos, desmintiendo también así aquella vinculación pretendidamente esencial y secular de la identificación de los términos «persona» y «ser humano».

E incluso en la modernidad, a través del llamado Derecho penal del enemigo, se pretende fundamentar la retirada de la condición de persona (y por ende la privación de derechos fundamentales) a ciertos sujetos, en concreto los terroristas, quienes al tener como finalidad destruir el ordenamiento jurídico, no quedan protegidos por él. Frente al Derecho penal del ciudadano, que sanciona y reinserta al que ha vulnerado el ordenamiento jurídico y la confianza social, se alza el Derecho penal del enemigo, que permite sancionar a quien pretende destruir aquel orden jurídico y social, aunque no lo haya vulnerado todavía, o imponerle una sanción mayor en caso de haberlo hecho y, en todo caso, darle un trato procesal con garantías disminuidas.

La fundamentación de esta concepción pasa por defender un concepto de persona en exclusiva formal, del tipo «Persona = ADN humano (o), ente racional/consciente/sintiente (o) y ente con función jurídica (1)». Así, Jakob Günter ha definido la persona como

[...] todo aquel que presta fidelidad al ordenamiento jurídico con cierta fiabilidad tiene derecho a ser tratado como persona, y quien no lleve a cabo esta prestación, será heteroadministrado, lo que significa que no será tratado como persona (Günter, 2005).

2.2.2. Embriones

Ya hemos apuntado que los defensores del modelo «Persona = ADN humano (I), ente racional/consciente/sintiente (o) y ente con función jurídica (o)» consideran, con independencia del reconocimiento del ordenamiento jurídico, que el ser humano es persona desde el mismo momento de la concepción. Así, argumentan que esta no comienza a existir después del hombre ni se extingue antes que él, haciendo de los conceptos vida humana y persona dos entes indisolubles. Como la persona (*alguien*) no puede derivar de una cosa (*algo*), el cigoto es persona, pues «el cigoto es distinto de la madre, no una parte de ella, con un programa genético propio y humano por lo que la nueva vida es, solo puede ser, humana» (Castiella, 2003: 19).

Ocurre que esta orientación choca frontalmente con la realidad: ningún ordenamiento jurídico moderno se decanta por la atribución de personalidad en el momento de la concepción, y en ninguna sociedad ni momento histórico se ha equiparado (en términos de duelo y consideración familiar y social) el fallecimiento de un niño de corta edad con la muerte de un embrión.

2.2.3. Animales

El esquema «Persona = ADN humano (o), ente racional/consciente/sintiente (I) y ente con función jurídica (o)», desde la perspectiva de aquellos autores que sitúan la capacidad biológica de sentir o la facultad de autoconsciencia como elemento nuclear del concepto, impone que la comunidad moral debe ser ampliada hasta incluir en ella a los animales que cumplan tales requisitos, lo que implicaría que el ordenamiento jurídico debería, en consonancia, considerarlos como sujetos de derechos.

Frente a esta pretensión, la mayoría de los ordenamientos jurídicos responden aún a la óptica tradicional procedente del Derecho romano, en la que el animal se concibe como un bien mueble semoviente y su estatuto jurídico responde a un criterio en exclusiva economicista. Esto se traduce en una regulación enfocada tan solo a animales objeto de explotación ganadera no industrializada o de caza, lo que no casa con las exigencias sociales y económicas de las modernas sociedades occidentales, en las que comienzan a quebrar aquellos principios para dar paso a la consideración de los animales como seres sensibles, en una línea iniciada por el Código austríaco en 1998 y que en España ha sido acogida por el Derecho. Dicha caracterización no supone, sin embargo, la creación de una nueva categoría en la dogmática jurídica, y desde esta perspec-

tiva siguen siendo objetos del derecho de propiedad, a los que supletoriamente se les aplica el régimen jurídico de las cosas muebles.

Y es que quizá la solución pase por una revisión de la dogmática tradicional y salvar la barrera que impide considerar a un ente sujeto de derecho y al mismo tiempo objeto de relaciones jurídicas. Se recuperaría así la sutileza del Derecho romano, que, aun negando al esclavo la condición de persona y, por tanto, la capacidad jurídica, le reconocía la capacidad de obrar (no la procesal) y le permitía la celebración de negocios jurídicos (recayendo sus efectos en la cabeza del dueño), para, por último, en la época imperial, sin alterar su *status*, reconocerle derecho a la vida, a la integridad corporal y moral.

2.3. EXCESOS

El modelo «Persona = ADN humano (o), ente racional/consciente/sintiente (o) y ente con función jurídica (I)» —y, en general, todos los esquemas propuestos en los que aparezca activado el elemento «ente con función jurídica»— da cabida a las concepciones de la persona que se alejan, por exceso de la equiparación de ese concepto, del ser humano de forma excluyente y exclusiva.

2.3.1. Personas jurídicas

El estudio histórico de la cuestión demuestra que, ya desde las primeras normas jurídicas, ciertos entes no humanos aparecen como destinatarios de aquellas: las referencias al palacio o a los dioses de los códigos de Hammurabi y Eshunna nos sitúan ante sujetos titulares de derechos de propiedad y de crédito por indemnizaciones por daños o robos. Y aunque, como es lógico, la dogmática de las personas jurídicas surge en un momento muy posterior, las leyes egipcias ya reconocieron aptitud legal para la titularidad de derechos a las fundaciones funerarias, al igual que hizo Roma con los *collegia*, a los que se les atribuye personalidad, no en atención al ejercicio del (entonces inexistente) derecho de asociación, sino por concesión del poder público. Y la historia también demuestra una progresiva distorsión, por ampliación, del concepto para incluir a partir de la influencia del Derecho canónico las *piae causae* como modelo de entidad jurídica sin base personal, a lo que se añade también toda la doctrina de las personas jurídicas de Derecho público y de Derecho internacional y la distrofia que supone todo el espectro de sociedades mercantiles, hasta reconocer las de tipo unipersonal.

Esta amalgama, con ingredientes de base personalista (asociaciones), pública (corporaciones), patrimonial (fundaciones) y comercial (sociedades), hace inútil cualquier intento descriptivo general: la persona jurídica no es en exclusiva un medio para aglutinar colectividades ni es el único medio jurídico existente para regular esos intereses pertenecientes a colectividades; no es tan solo un medio para limitar responsabilidades patrimoniales, ni siempre lo hace ni es la única solución jurídica para limitar la responsabilidad patrimonial; tampoco persigue siempre la obtención de un beneficio económico, y existen múltiples vías jurídicas para la consecución de beneficio económico.

No es de extrañar que todas las teorías que han intentado ofrecer una visión unitaria del fenómeno hayan fracasado, salvo las que lo hacen desde una óptica exclusivamente formal y funcional. Ferrara (2006: 93-268), desbrozando las diversas teorías que han tratado de explicar este fenómeno, explica cómo deben rechazarse las que defienden que la persona jurídica es una mera ficción, pues la persona jurídica responde a una realidad preexistente. Y no siempre se tratará de un patrimonio vinculado a un fin (Brinz) ni siempre será una persona colectiva (Thöl y Gierke). Tampoco son todos organismos sociales (Lilienfeld) ni voluntades incorpóreas (Zitelmann) más allá de en una relación de analogía vaga y referencial. Para este autor, la personalidad jurídica es una forma, no un ente. En la vida social existirán colectividades de seres humanos con finalidades o negocios comunes, o cotitulares de unos mismos bienes, o patrimonios que por voluntad de su dueño se aplican a cierta finalidad, pero nada de ello permite hablar de personas en potencia o personas de hecho, sino, a lo sumo, de aspirantes a la personalidad, a la que solo acceden cuando cumplen los requisitos que para ello impone el ordenamiento jurídico.

2.3.2. Personas electrónicas

El estudio de la (eventual) personalidad jurídica de los sistemas de inteligencia artificial es objeto de este trabajo. En la práctica se ha impuesto la denominación «personalidad jurídica electrónica», terminología empleada también por las instituciones de la Unión Europea. No obstante, puede ser demasiado restrictiva ante la posibilidad de sistemas de computación de base molecular, enzimática o proteica —biológica, en definitiva—, o de naturaleza cuántica. Por ello, en este trabajo se ha optado por utilizar también la expresión «personalidad algorítmica», en alusión a su sede generadora sin prejuzgar la naturaleza del sistema que la contiene.

3.1. ROBOTS E INTELIGENCIA ARTIFICIAL

3.1.1. Del mito a la realidad

El sueño de crear androides con características humanas ha sido una constante en todas las culturas. En la historia encontramos diversos ejemplos, como las estatuas de oro ayudantes de Hefesto de las que habla Homero en la *Iliada*; Talos, el hombre de bronce mencionado por Apolonio de Rodas en las *Argonáuticas*; el mito de Pigmalión; el Gólem de Praga o los Robots Universales de Rossum en la obra teatral de Karel Čapek. Junto a los mitos y la literatura encontramos también ejemplos históricos, o bien con una pretensión solo lúdica, como los *karakuri ningyō* —autómatas japoneses contruidos en los siglos XVIII y XIX— o el Turco —un artilugio experto en ajedrez construido por Wolfgang von Kempelen en 1769—, o bien con una finalidad de trascendencia, como el *Ars magna* de Ramon Llull —una máquina que se proponía demostrar la veracidad de la fe cristiana mediante la combinación de cincuenta y cuatro ideas simples o raíces.

Remedo de los humanos, tales creaciones fueron también investidas idealmente de las mismas virtudes de aquellos: las estatuas de oro de la *Iliada* estaban dotadas de inteligencia; en Talos, el hombre de bronce, latían «implicaciones para la vida y la muerte» y el Turco asombró a Europa y América con su precisión técnica, venciendo a los mejores jugadores de ajedrez de su tiempo, aunque para ello se sirviese de otra característica moral propia de los humanos como es la mentira, pues no era un ingenioso mecanismo de relojería el que decidía los magistrales movimientos, sino que, en su interior, se ocultaba William Schlumberger, uno de los mejores jugadores de ajedrez de aquellos años.

En el tratamiento del aspecto ético de estos artificios tecnológicos, el mundo literario se adelantó apenas dos décadas al mundo real. Mientras Isaac Asimov escribía sus leyes de la robótica en 1942 en su relato *Círculo vicioso*, en el que condensaba el código moral de los androides en «formulaciones matemá-

ticas impresas en los senderos positrónicos del cerebro», la robótica real aún estaba muy lejos del ámbito moral: su máximo logro por aquella época fueron dos tortugas metálicas, Elmer y Elsie, diseñadas en 1948 por William Grey Walter, que, provistas de sensores lumínicos y táctiles, conseguían evitar obstáculos en una habitación, ocultarse de la luz e interactuar entre sí.

No obstante, ya en esos años, en los tribunales de Estados Unidos se plantearon casos relativos a la responsabilidad por daños ocasionados por robots. En el caso *Brose vs. Estados Unidos*,

[...] un tribunal federal de distrito tuvo que decidir si el gobierno estadounidense podría ser considerado responsable de un accidente de avión entre un avión de combate y un pequeño avión privado; a juicio del Tribunal estaba claramente demostrado que el avión del ejército en el momento de la colisión estaba en modo *robot-control* (Calo, 2016: 24).

Es indiscutible que el grado de autonomía de aquellos robots primitivos era incomparable (y no solo en lo cuantitativo) al nivel alcanzado en la actualidad, pero la mención expresa al concepto es relevante.

En el año 1956, confluyen dos factores decisivos para el desarrollo tecnológico: la compañía Unimation, fundada por George Devol y Joseph Engelberger, presenta el primer robot comercial, Unimate, un brazo articulado programado para transportar material fundido y soldar piezas en una cadena de montaje de automóviles, y John McCarthy acuña el término «inteligencia artificial» para referirse a «la ciencia e ingenio de hacer máquinas inteligentes, especialmente programas de cómputo inteligentes». La robótica, en la actualidad, es una rama de la ingeniería que se ocupa del estudio, diseño, fabricación y empleo de los robots, y abarca una diversidad de disciplinas, entre ellas la inteligencia artificial. A su vez, esta se presenta también como un concepto autónomo y multidisciplinar que implica distintas materias (informática, filosofía, biología y matemáticas), y estudia la creación y el diseño de sistemas capaces de resolver problemas cotidianos por sí mismos utilizando como modelo de referencia la inteligencia humana.

Los expertos en esta materia afirman que los límites de la inteligencia artificial como rama científica son muy difusos, pues en su desarrollo intervienen disciplinas como la programación semántica, la lógica, la neurociencia, la computación, las matemáticas, la biología y la estadística. En cualquier caso, el desarrollo de la robótica y de la inteligencia artificial y la creación de máquinas dotadas de autonomía y capacidad de decisión anuncia, de forma inminente, una nueva revolución industrial. A pesar de que aún están en ciernes, estos

sectores han triplicado en el último decenio las solicitudes de patentes y han aumentado sus ventas una media del 17 % anual en los cinco primeros años de la presente década, según el *Proyecto de informe con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica* de la Comisión de Asuntos Jurídicos del Parlamento Europeo del 31 de mayo de 2016.

Pero más llamativo y relevante, desde una perspectiva bioética, que el aspecto cuantitativo del desarrollo de la industria y el comercio de la robótica y la inteligencia artificial es el cambio cualitativo radical que se está produciendo en el modo de relacionarse el ser humano con las máquinas. En efecto, «en un grado mayor que cualquier otra tecnología en la historia, los robots tienen un impacto o valor social para las personas» (Barrio y García-Prieto, 2018: 25). Y este mayor impacto social se explica por varias razones. En primer lugar, por un motivo puramente psicológico, pues los experimentos realizados por Peter Khan —y citados por aquel— llevan a una

[...] sorprendente conclusión: los robots pueden pertenecer a una categoría ontológica completamente nueva. Los participantes no tienden a pensar que los robots personificados están vivos, pero tampoco los consideran objetos. Más bien, los partícipes en esos estudios están inclinados a atribuir estados mentales a los robots, e incluso adoptan comportamientos que serían impensables al tratar con un mero objeto (Barrio y García-Prieto, 2018: 25).

Claro que, frente a esta razón perceptiva, existen otras de carácter objetivo que explican esa transformación en la relación ser humano-máquina, pues si bien en una primera fase la seguridad e integridad física de los usuarios era el único elemento a proteger de los eventuales riesgos generados por la implantación de los robots, la mayor complejidad y autonomía de estos, así como la exponencial ampliación de los ámbitos en los que desempeñan sus funciones, han generado la necesidad de establecer medidas jurídicas de protección que abarcan otros muchos aspectos. De este modo, la intimidad personal y familiar, la salud, la dignidad, la autonomía, la propiedad y la protección de los datos de carácter personal son elementos esenciales en la regulación actual de los derechos humanos, por lo que merecen una especial atención. En todos esos aspectos incide el *Proyecto de informe con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica* de la Comisión de Asuntos Jurídicos del Parlamento Europeo del 31 de mayo de 2016 y la posterior Resolución del Parlamento Europeo.

3.1.2. De qué hablamos cuando hablamos de robots

Según la documentación de la Robot Industries Association (el conglomerado empresarial más importante del mundo sobre robótica), el robot puede definirse como una máquina funcional reprogramable, capaz de mover material, piezas, herramientas o dispositivos especializados mediante movimientos variables programados con el fin de realizar tareas diversas y caracterizada por su elevado grado de autonomía.

Aunque esta definición debe ser completada, especialmente desde el punto de vista funcional en atención a la diversidad de papeles que desempeñan en la actualidad, tiene el mérito de destacar una de las características esenciales del robot: su autonomía; esencial, sí, pero relativa y susceptible a una amplia gradación. Así, la autonomía implica, por significado (la condición de no depender de nadie) y por etimología (*auto*, yo; *nomos*, ley), una llamada al autogobierno, y no es un concepto estático y predeterminado, sino que ha de referirse a objetivos concretos. Circunscribiendo el concepto al ámbito tecnológico, un sistema se califica de autónomo cuando puede actuar de acuerdo con sus propios objetivos, normas internas, preceptos y conocimiento, y sin intervención exterior (Williams, y Scharre, 2015: 13-15). Un sistema autónomo es, pues, capaz de decidir un curso de acción a partir de una serie de alternativas sin depender de la supervisión y el control humanos, aunque estos todavía pueden estar presentes.

Más allá de la mera automatización, la autonomía implica la facultad de adaptación a entornos cambiantes y la solución de problemas a partir de planteamientos incompletos o en entornos con (mayor o menor) dosis de incertidumbre. Y así, superando las respuestas automáticas del termostato para mantener la temperatura de una habitación en los valores prefijados, un sistema autónomo de detección de objetivos despegue, vuela con condiciones atmosféricas cambiantes, localiza, fija los blancos, comunica su situación, regresa a la base y aterriza al finalizar su misión o, en caso de accidente, pérdida de energía o captura por parte del enemigo, se autodestruye.

Con respecto a su autonomía, se distinguen varias generaciones de robots. La primera se corresponde con máquinas con capacidad para almacenar trayectorias programables de movimiento repetitivo descritas punto a punto y dotadas de sensores internos, cuyo ejemplo paradigmático es el brazo manipulador de las cadenas de montaje. A finales de los años setenta, surgen los robots adaptativos, que encarnan la segunda generación robótica. Son robots que disponen de sensores externos (temperatura, tacto y visión), que les suministran información del mundo exterior y que permiten una retroalimentación que los

faculta para hacer elecciones limitadas y reaccionar ante cambios en las circunstancias exteriores. En la tercera y actual generación, desempeña un papel esencial la inteligencia artificial, que permite a los robots alcanzar lo que técnicamente se denomina razonamiento simbólico: ya no se trata de trabajar con algoritmos basándose en información exterior obtenida a través de sensores, sino de extraer reglas a partir de principios y generalizar su aplicación a situaciones antes desconocidas, así como tolerar excepciones a reglas predeterminadas o aprendidas (Barrio y García-Prieto, 2018: 6).

Ya se han mencionado las diferentes funciones que pueden desempeñar los robots. Esa funcionalidad nos lleva a una primera clasificación de los robots:

- a) Industriales: la fundición fue el primer sector en el que se implantaron sistemas robóticos, que en la actualidad están extendidos a casi todos los ámbitos de la actividad productiva: cadenas de montaje, paletizado y empaquetado, extracción, modelado, control de calidad, ingeniería inversa, agricultura, ganadería, manipulación, etc.
- b) Sanitarios: compendio de la complejidad técnica y de la cercanía humana, el sanitario es uno de los ámbitos en los que la generalización de los robots plantea mayores dilemas éticos; sus aplicaciones concretas abarcan desde el diagnóstico a la cirugía, pasando por la organización logística hospitalaria, la rehabilitación, la asistencia y el cuidado personal, etc.
- c) Militar y policial: el 8 de julio de 2016 un robot Remotec Modelo F-5 fue programado por la policía de Dallas para, mediante explosivos, acabar con la vida de un sospechoso que se encontraba atrincherado tras matar a seis policías. Esto abrió un debate que ya ha llegado a la agenda de la ONU (Ginebra, noviembre de 2017).
- d) Transporte: el sector de los automóviles autónomos, por su desarrollo y potencial de comercialización, reclama con urgencia una normativa de alcance internacional, al igual que ocurre con los vehículos aéreos no tripulados.
- e) Recreativo y educativo: la robótica está llamada, a la luz de las investigaciones y prototipos actuales, a generalizarse en una diversidad de ámbitos: compañía, juego infantil y de adultos, sexualidad, enseñanza teórica y práctica, etc.

Esta clasificación, al igual que el propio concepto propuesto antes, no alcanza a poner de relieve la variedad morfológica y estructural de los robots. Al alejarse de la visión estereotipada del robot humanoide (o, en el extremo opuesto, del brazo mecánico de la cadena de montaje), esta categoría abarca un am-

plio espectro tecnológico, en el que se incluyen tanto artefactos de tipo hidráulico-mecánico, como componentes nanotecnológicos (nanorrobots, máquinas moleculares), biorrobots, prótesis, *soft-bots* y sistemas de inteligencia artificial.

3.2. ROBOÉTICA Y DERECHO DE LOS ROBOTS

En 1987, M. Mitchell Waldrop publica en *The AI Magazine* un artículo titulado «Una cuestión de responsabilidad». En él, perfila el concepto de «ética de las máquinas» y propone la necesidad de controlar los valores, suposiciones y propósitos introducidos por programadores para concretar aquella capacidad de decisión.

En el año 2002, Gianmarco Veruggio acuña el término «roboética», que define como una parte de la ética aplicada que se ocupa del estudio de los problemas derivados de la interacción de los sistemas de inteligencia artificial con el ser humano, los animales y la naturaleza, y tiene como objetivo el desarrollo de herramientas técnico-científicas y culturales que promuevan la robótica como causa de avance de la sociedad humana y de sus individuos, y que ayuden a prevenir un uso equivocado de aquella contra la propia especie humana.

Los ámbitos en los que pueden plantearse esos problemas éticos son principalmente tres:

- 1) La fabricación de los robots y la programación de los sistemas de inteligencia artificial, donde se debe garantizar el respeto a la dignidad, intimidad y seguridad de las personas.
- 2) La realización de funciones por robots y sistemas de inteligencia artificial que impliquen la toma de decisiones de carácter ético y actuaciones autónomas e impredecibles, con la posible consideración de aquellos como agentes morales y la eventual problemática de la determinación de la responsabilidad por daños generados por tales sistemas.
- 3) El tratamiento que los ordenamientos jurídicos deben dar a los sistemas de inteligencia artificial y su posible consideración como pacientes morales o directamente como sujetos de derecho.

La forma peculiar en que el ser humano percibe a los robots y a los sistemas de inteligencia artificial, la variedad de los derechos e intereses jurídicos implicados y su relevancia para una amplia gama de contextos jurídicos (derecho de la persona, responsabilidad civil y penal, propiedad industrial e intelectual,

contratos, navegación aérea, marítima y terrestre, tributos, seguridad social, etc.; Barrio y García-Prieto 2018: 26-27) han llevado a plantear incluso la viabilidad del Derecho robótico como una disciplina específica. Y a juicio de Moisés Barrio Andrés (2018), se trata más de necesidad que de mera viabilidad, pues resulta apremiante la construcción de un Derecho de los robots que pueda funcionar como marco jurídico e institucional para lograr un equilibrio entre los derechos fundamentales y las consecuencias de la comercialización, expansión y desarrollo de los sistemas robóticos y de inteligencia artificial.

3.3. LA SITUACIÓN LEGAL ACTUAL

3.3.1. *Lege data*

Aunque experimentos psicológicos como los citados muestran que la percepción que los seres humanos tienen de los robots y de los sistemas de inteligencia artificial los sitúa en un campo distinto del de las meras cosas, esta percepción es ajena a la realidad jurídica actual. Los robots y los sistemas de inteligencia artificial existentes hoy en día son, respectivamente, máquinas y *software*. Por lo tanto, en sentido jurídico, se consideran cosas y, como tales, susceptibles de ser objeto, pero no sujeto de derechos.

Al carecer de personalidad jurídica, todas las consecuencias que deriven de los actos realizados por los robots o sistemas de inteligencia artificial recaerán de manera directa sobre las personas (físicas o jurídicas) que, según las circunstancias, corresponda según la ley en virtud del vínculo jurídico que lo ligue con aquellos (propietario, arrendatario, licenciatario de *software*, usufructuario, etc.). Y tales consecuencias jurídicas pueden derivar tanto de contratos como de actos de naturaleza extracontractual.

Entre los primeros puede citarse la contratación algorítmica de alta frecuencia (a la que se refiere la Directiva UE 2014/65 del 15 de mayo). Esta se define como la negociación de instrumentos financieros en la que un algoritmo informático determina automáticamente los distintos parámetros de las órdenes (si la orden va a ejecutarse o no, el momento, el precio, la cantidad, cómo va a gestionarse después de su presentación), con una intervención humana limitada o nula. La normativa reguladora de estas operaciones trata de mitigar sus efectos perturbadores sobre los mercados, pues son fuente frecuente de alteración de su funcionamiento normal y elevan el riesgo sistémico y la volatilidad, incluso en los casos en los que no se utilizan con una finalidad fraudulenta. *Spoofing*, *layering*, *quote stuffing* son términos todos ellos que designan prácticas prohibi-

das, pero el hecho de lanzar y cancelar órdenes para despistar a otros negociadores o para aprovechar la tendencia así creada (a eso se refieren esos vocablos anglosajones) puede pasar con facilidad desapercibido cuando el número de órdenes cruzadas se aproxima a las cuarenta mil cada diez microsegundos.

La velocidad de análisis y respuesta, el tamaño de las bases de datos manejadas y los grandes números (las variaciones ínfimas en oferta y demanda en millones de operaciones cruzadas se traducen en ganancias o pérdidas millonarias) son las claves de estos sistemas, que se presentan como el paradigma de las potencialidades que pueden ofrecer en el ámbito de la contratación en masa.

Pero es, sin duda, en el ámbito de la responsabilidad extracontractual donde con mayor urgencia se reclama la necesidad de una nueva regulación, ya que los viejos esquemas de la responsabilidad por culpa no se adaptan a las necesidades generadas por el alto nivel de autonomía de los robots.

En efecto, el artículo 1902 de nuestro Código civil establece que «el que por acción u omisión causa daño a otro, interviniendo culpa o negligencia, está obligado a reparar el daño causado». Cuando se trata de daños causados por una máquina, un robot o un sistema de *software*, es una cuestión técnico-jurídica determinar si el responsable civil ha sido el propietario, el operario o el fabricante. El tenor literal de la norma citada, cuya aplicación exige la concurrencia de culpa en el sujeto causante del daño, ha sido atenuado por la jurisprudencia mediante inversiones en la carga de la prueba o exigencias de determinados grados de culpa, incluso sin concurrencia de conductas antijurídicas, de manera que nuestro ordenamiento conduce a un sistema de tipo objetivo en el que lo esencial es la producción de un daño efectivo a un tercero a causa del riesgo generado. Algunas normas de nuestro decimonónico Código civil ya responden a esta tendencia moderna (por ejemplo, los artículos 1905, 1908 y 1909 respecto a daños causados por animales, explosiones de maquinaria o defectos constructivos), y determinados sectores se encuentran, en virtud de normas especiales, sujetos a un régimen de responsabilidad por riesgo (por ejemplo, respecto a daños personales en accidentes de circulación) o, directamente, objetiva (como en materia de navegación aérea o accidentes nucleares).

Un sistema de responsabilidad objetiva es también el diseñado por la Directiva 85/374/CEE del Consejo, del 25 de julio de 1985, relativo a la aproximación de las disposiciones legales, reglamentarias y administrativas de los Estados miembros en materia de responsabilidad por los daños causados por productos defectuosos. Esta norma, ya traspuesta al ordenamiento español mediante la legislación protectora de consumidores y usuarios (actualmente, Real Decreto Legislativo 1/2007, del 16 de noviembre que aprueba el Texto Refundido de la Ley General para la Defensa de los Consumidores y Usuarios), resulta, sin

embargo, insuficiente para cubrir los daños causados por la nueva generación de robots. Esto se debe a que se los puede dotar de capacidades de adaptación y aprendizaje que entrañan cierto grado de imprevisibilidad en su comportamiento, ya que podrían aprender de forma autónoma de su propia experiencia variable e interactuar con su entorno de forma única e imprevisible, según afirma el *Proyecto de informe con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica* de la Comisión de Asuntos Jurídicos del Parlamento Europeo del 31 de mayo de 2016.

3.3.2. *Lege ferenda*

Impulsado y financiado por la Unión Europea, el proyecto Robo-Law, desarrollado entre el 1 de marzo de 2012 y el 22 de septiembre de 2014, perfiló las líneas más importantes a tener en cuenta en una futura normativa sobre robótica, y trata de dar respuestas a tres cuestiones básicas: las herramientas legales que mejor se adaptan al objetivo de regular la tecnología robótica y la inteligencia artificial, el tipo de análisis ético que se debe realizar sobre la innovación tecnológica y la metodología a aplicar, respetando la política de la Unión Europea sobre Investigación e Innovación Responsable (RRI) y los contenidos de la regulación, con especial atención a la necesidad de proteger los derechos fundamentales de los ciudadanos europeos.

Superando el interés puramente doctrinal y con una finalidad práctica, el 3 de septiembre de 2013, la Comisión Europea, junto con un consorcio de diez empresas europeas, puso en marcha el proyecto PETROBOT con el objetivo de fabricar robots que puedan sustituir, ejerciendo sus mismas capacidades, a los seres humanos en las inspecciones de los recipientes a presión y los tanques de almacenamiento que se utilizan de forma generalizada en los sectores del petróleo y el gas y en la industria petroquímica.

Solo un mes después, el eurodiputado Marc Tarabella formuló al Parlamento Europeo una cuestión escrita en la que solicitaba a la cámara, en relación con esa eventual sustitución de personal humano por robots, que se pronunciase acerca del estatuto jurídico de los robots y de su posición sobre un eventual reconocimiento de su personalidad jurídica. En su respuesta del 15 de noviembre de 2013, la Comisión afirmó que

[...] la tecnología aún no está lista para ofrecer el grado de autonomía necesario para otorgar personalidad o estatus legal a los robots. Sin embargo, la comunidad de la robótica y de los sistemas cognitivos artificiales tiene muy en cuenta las

cuestiones éticas, jurídicas y sociales relacionadas con los sistemas futuros, dotados de mayor inteligencia y autonomía.

Asimismo, afirma que aunque los sistemas totalmente autónomos aún no son una realidad, existen tendencias mundiales claras que indican que podríamos ver tales sistemas en el futuro, concluyendo que

[...] el desarrollo de opciones políticas y la comprensión de las consecuencias jurídicas de los sistemas plenamente autónomos es una forma de preparar el futuro y que la concesión de personalidad jurídica a los robots es actualmente un debate académico.

Con estos antecedentes, el 31 de mayo de 2016, la Comisión de Asuntos Jurídicos del Parlamento Europeo aprobó el *Proyecto de informe con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica*, cuyas líneas esenciales se resumen a continuación.

Con referencias expresas a figuras literarias y mitológicas, como el monstruo de Frankenstein creado por Mary Shelley, Pigmalión y el Gólem de Praga, el proyecto normativo parte del expreso reconocimiento de que la humanidad se encuentra a las puertas de una era en la que robots, bots, androides y otras formas de inteligencia artificial cada vez más sofisticadas van a desencadenar una nueva revolución industrial, que es probable que afecte a todos los estratos de la sociedad.

En la introducción del Proyecto se ponderan los eventuales beneficios de los sistemas robóticos y de inteligencia artificial con sus riesgos. En el primer aspecto, se reseñan cuestiones tales como la eficiencia, el ahorro y la seguridad humana en sectores como la producción, el comercio, el transporte, la asistencia sanitaria, la educación y la agricultura. El contrapeso lo sitúa en los riesgos, más que probables, que su generalización supondría para:

- el desempleo y la viabilidad de los sistemas de seguridad social;
- la seguridad física, en caso de fallos de programación o de ataques informáticos contra robots interconectados y sistemas robóticos;
- la protección de los datos personales y la privacidad;
- la dignidad humana, en especial respecto de los robots que ejerzan funciones de cuidados, compañía, rehabilitación e intervenciones sanitarias.

El Proyecto de informe sobre robótica parte del reconocimiento expreso de que, en el estado actual de la técnica, los robots y los sistemas de inteligencia

artificial no tienen consciencia de su propia existencia, pero con independencia de ello, es procedente regular las cuestiones relativas a la responsabilidad civil y la deontología, tomando en consideración los valores intrínsecamente europeos y humanistas que caracterizan la contribución de Europa a la sociedad y adoptando un enfoque gradual, prudente y pragmático.

La cuestión de la responsabilidad jurídica derivada de la actuación perjudicial de un robot es, para el proyecto normativo europeo, una cuestión crucial, dado el grado de autonomía que han alcanzado. En efecto, a mayor autonomía, menor es la posibilidad de considerarlos simples instrumentos en manos de otros agentes (fabricante, propietario, usuario), lo que hace que la normativa general sobre responsabilidad resulte insuficiente, tal y como ya se ha expuesto al citar la Directiva 85/374, que solo cubriría los daños ocasionados por un defecto, demostrado, de fabricación.

Asimismo, el proyecto normativo destaca las deficiencias del marco jurídico actual en el ámbito de la responsabilidad contractual ante la proliferación de máquinas concebidas para elegir a sus contrapartes, negociar cláusulas, precios y formas de pago, celebrar contratos y decidir su ejecución.

Las razones que determinan la insuficiencia del marco normativo actual se explican con claridad en el Proyecto:

- Por un lado, la autonomía de los robots, que define como la capacidad de tomar decisiones y aplicarlas en el mundo exterior, con independencia de cualquier control o influencia externa.
- Por otro, la imprevisibilidad de su funcionamiento, derivada de la capacidad de adaptación y aprendizaje, que permite a las nuevas generaciones de robots interactuar con su entorno de forma única y no predeterminada.

El Proyecto reconoce que, en el actual marco jurídico, los robots no pueden, en sí mismos, ser considerados responsables de los actos u omisiones que causan daño a terceros. Esto se debe a que, tal y como se ha expuesto, las normas vigentes en materia de responsabilidad abarcan aquellos casos en los que es posible remontarse hasta un agente humano concreto, como el fabricante, el propietario o el usuario. A estos agentes debe también añadirse, en el caso de robots con capacidad de aprendizaje y autonomía, el «profesor» encargado del «entrenamiento» o la «educación», teniendo muy en cuenta que las competencias adquiridas a través de esa «educación» no deben confundirse con las competencias estrictamente dependientes de su capacidad de aprender de modo autónomo.

Las ideas en torno a las cuales el proyecto normativo propone una nueva regulación de esta cuestión esencial de la responsabilidad civil son las siguientes:

- No limitación, en modo alguno, del tipo o el alcance de los daños y perjuicios que puedan ser objeto de compensación, ni tampoco limitar la naturaleza de dicha compensación por el único motivo de que los daños hayan sido causados por un agente no perteneciente a la especie humana.
- El modelo normativo a seguir debería basarse en un sistema de responsabilidad objetiva, lo que tan solo exige probar que se ha producido un daño y el establecimiento de un nexo causal entre el comportamiento perjudicial del robot y los daños causados.
- Conveniencia del establecimiento de un régimen de seguro obligatorio a cargo de los fabricantes de robots autónomos y completar dicho sistema con un fondo de compensación para garantizar la reparación de daños en los casos de ausencia de una cobertura de seguro y permitir, además, otras operaciones financieras en interés del robot, tales como inversiones, donaciones o remuneraciones por servicios, que podrían transferirse al fondo.
- El citado fondo de compensación permitiría establecer un régimen de responsabilidad limitada para el fabricante, el programador, el propietario o el usuario de los robots inteligentes autónomos, de modo que solo podrían invocarse los daños materiales dentro de los límites de dicho fondo. Otros tipos de daños no estarían sujetos a tales límites.
- Necesidad de abordar la cuestión fundamental de si los robots deben tener personalidad jurídica, pues la citada autonomía de los robots suscita la cuestión de su condición y pertenencia a una de las categorías jurídicas existentes —es decir, si deben considerarse personas físicas, personas jurídicas, animales u objetos— o la creación de una nueva categoría, como la de personas electrónicas, especialmente en los supuestos en los que los robots puedan tomar decisiones autónomas inteligentes o interactuar con terceros de forma independiente.

El resultado práctico del Proyecto de informe fue la Resolución del Parlamento Europeo del 16 de febrero de 2017, que, en lo esencial, sigue el esquema y reproduce el contenido del Proyecto de informe, aunque de forma expresa precisa que la atribución de personalidad jurídica a los robots y sistemas de inteligencia artificial se plantea a largo plazo

[...] de forma que como mínimo los robots autónomos más complejos puedan ser considerados personas electrónicas responsables de reparar los daños que puedan causar, y posiblemente aplicar la personalidad electrónica a aquellos supuestos en los que los robots tomen decisiones autónomas inteligentes o interactúen con terceros de forma independiente.

Esta última propuesta, la atribución de personalidad jurídica a los robots, enlaza de forma directa e inmediata con el objetivo de este trabajo, por lo que debemos plantearnos cuáles son las razones favorables y contrarias a ese eventual reconocimiento.

3.4. ¿POR QUÉ PERSONAS? PROS Y CONTRAS

En el ámbito subjetivo de la comunidad moral, podemos distinguir: al mero espectador, sujeto moral que no es actor de la acción en cuestión pero presencia la acción del sujeto y la comprende como moralmente relevante; al paciente moral, destinatario de la conducta del agente moral, y al agente moral, concebido como un sujeto autónomo, «capaz de evaluación y decisión moral» (Rodríguez, 2012: 127-138) y responsable de las consecuencias de sus acciones.

3.4.1. El agente moral

Ya se ha mencionado el caso de la contratación algorítmica de alta frecuencia: complejos algoritmos analizan, con nula o limitada intervención humana, situaciones de los mercados financieros, fijan precios y condiciones, concertan y ejecutan contratos de instrumentos financieros. Y más allá de ese ámbito de la contratación bursátil, en el marco de la actual economía del *big data*, sistemas informáticos procesan miles de currículos o solicitudes de préstamos en un par de segundos y los clasifican en listas bien ordenadas con los candidatos más adecuados situados en los primeros puestos.

Como un mecanismo elemental de salvaguardia de los derechos de los ciudadanos frente a tales sistemas, el artículo 13 de la Ley Orgánica 15/1999 del 13 de diciembre de Protección de Datos de Carácter Personal establece que los ciudadanos tienen derecho a no verse sometidos a una decisión con efectos jurídicos que se base tan solo en un tratamiento de datos. Asimismo, se determina que la valoración sobre el comportamiento de los ciudadanos basada en un tratamiento de datos únicamente podrá tener valor probatorio a petición

del afectado. En un sentido similar se pronuncian tanto el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo del 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, y la Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo del 27 de abril de 2016 relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales.

Otros ejemplos de máquinas que adoptan decisiones con trascendencia moral y jurídica los encontramos en los vehículos (terrestres y aéreos) no tripulados, así como en los robots asistenciales, de diagnóstico, quirúrgicos y protésicos. Su autonomía y complejidad aumentan día a día y, en la medida en que conducen a la imprevisibilidad de su comportamiento, con la consiguiente imposibilidad de atribución de responsabilidad a un agente humano, nos sitúan a las puertas de lo que el Proyecto de informe y la Resolución del Parlamento de la Unión Europea catalogan como insuficiencia del sistema normativo vigente.

Debemos, pues, plantearnos si el hecho de que los robots y sistemas de inteligencia artificial ejecuten decisiones con consecuencias morales o jurídicas los transforma de inmediato en agentes morales y, en caso afirmativo, si esa condición de agente moral conduce de forma necesaria a su catalogación como sujetos de derecho.

3.4.2. Agentes morales implícitos, explícitos y plenos

Profundizando en la posible relación de los sistemas computacionales con la agencia moral, se distingue entre agentes morales implícitos, agentes morales explícitos y agentes morales plenos (Moor, 2006: 18-21).

Según este autor, pueden catalogarse como agentes morales implícitos aquellos sistemas computacionales que no ejecutan decisiones éticas, pero que por diseño y programación tienen restringidas sus acciones para evitar resultados no éticos. En lugar de programar un *software* con máximas éticas explícitas entre las que el sistema deba elegir en cada caso, estos sistemas actúan éticamente porque sus funciones internas así lo imponen: los cajeros automáticos o los sistemas de banca *online* serían un ejemplo de sistemas éticos implícitos que no constituyen un agente moral pleno, pero cuyas funciones obligan a sus diseñadores a considerar de manera rutinaria la ética en la programación, impidiendo, por ejemplo, el acceso a cuentas ajenas al usuario.

Los agentes morales explícitos se corresponden con sistemas computacionales que emiten juicios y toman decisiones de carácter ético. Tales juicios y decisiones responden a los criterios éticos que sus programadores han incluido en su código fuente, de modo que se limitan a cumplir las normas, los criterios o los valores morales con los que han sido programados en los contextos en los que se utilizan. Son ejemplos los ya citados sistemas informáticos de calificación de solvencia bancaria y de currículums laborales, o el formulario LSI-R utilizado por la mayoría de los estados de Estados Unidos para valorar el riesgo de reincidencia de la población reclusa en la fijación de condenas y concesión de la libertad condicional. Es evidente que estos sistemas automatizados de valoración de riesgos se limitan a aplicar los mismos criterios éticos que los desarrolladores (con sus eventuales sesgos y prejuicios) han incluido en su programación y, por lo tanto, aunque formulen o ejecuten juicios éticos de forma explícita, tampoco actúan como verdaderos agentes morales, pues lo hacen «en nombre» de sus programadores o usuarios.

Los agentes morales plenos no solo emiten y ejecutan juicios éticos de forma explícita, sino que también son competentes para justificarlos de un modo razonado. El ser humano adulto promedio es el prototipo de lo que se considera un agente moral pleno. No obstante, en el estado actual de la tecnología, esa cualidad —la competencia para justificar decisiones éticas— no es atribuible a los sistemas de inteligencia artificial, aunque lo cierto es que no puede afirmarse que no lleguen a disponer de ella en el futuro. Imaginemos el hipotético supuesto de dos asistentes hospitalarios de cuidados que, ante una subida de la fiebre de sus respectivos pacientes, optan, en un caso, por avisar al médico supervisor y, en el otro, por aplicar el antipirético pautado para tales casos por el historial clínico. Imaginemos que el primero de los pacientes fallece y el otro se salva. Parece claro que ambas opciones son moralmente calificables y atribuibles a agentes morales y, ante un supuesto así, cabe plantearse si el hecho de que uno de los cuidadores fuera humano y el otro un robot de cuidados autónomo alteraría la conceptualización de sus actos como cualificados a nivel moral y procedentes de un agente moral (Floridi y Sanders, 2004: 1-21).

3.4.3. Características del agente moral: autonomía, interactividad y adaptabilidad

Floridi y Sanders (2004) articulan la noción de agente moral sobre tres pilares: interactividad, autonomía y adaptabilidad. Podemos calificarlos como esencia-

les, pues son un mínimo necesario y, si falta cualquiera de ellos, el concepto de agente moral se derrumba:

- 1) La interactividad significa que el agente y su entorno pueden actuar recíprocamente.
- 2) La autonomía implica una actuación independiente, se concibe como la capacidad de elegir, de hacer y actuar según las propias decisiones, y significa que el agente es capaz de cambiar de estado sin que ello suponga una respuesta automatizada a la interacción.
- 3) La adaptabilidad es la aptitud para adecuarse a las condiciones del entorno. Significa que, ante alteraciones en este último, las interacciones del agente pueden cambiar las reglas de transición por las cuales modifica su estado.

Para ilustrar estas nociones, los autores citados exponen algunos ejemplos de sistemas en los que concurren todas o algunas de estas notas y que podemos articular en las siguientes combinaciones atendiendo a su presencia o no (de nuevo se señala con (I) la presencia del elemento y con (o) su ausencia):

- a) Interactividad (o); autonomía (o); adaptabilidad (o): se trata de sistemas por completo inertes, como, por ejemplo, las rocas, respecto de los cuales solo en un sentido muy amplio e impropio podríamos hablar de cierta interactividad; por ejemplo, en atención a su erosión por el paso del tiempo.
- b) Interactividad (I); autonomía (o); adaptabilidad (o): son sistemas que reaccionan de modo automático ante ciertas alteraciones del entorno, como, por ejemplo, un molino de viento.
- c) Interactividad (o); autonomía (I); adaptabilidad (o): son sistemas que actúan con independencia del entorno o, al menos, mientras no se alteren las condiciones de este, como, por ejemplo, a un bajo nivel, un péndulo.
- d) Interactividad (o); autonomía (o); adaptabilidad (I): los propios Floridi y Sanders reconocen que ni siquiera en un sentido vago y general podría encontrarse un sistema en el que concurriese solo la nota de la adaptabilidad sin interactividad ni autonomía.
- e) Interactividad (o); autonomía (I); adaptabilidad (I): se citan los ejemplos de los sistemas planetarios y de ecosistemas cerrados.
- f) Interactividad (I); autonomía (o); adaptabilidad (I): son sistemas sensibles a los cambios del entorno, a los que responden de forma automatizada. Aquí se incluyen los termostatos o los sistemas tradicionales de pilotaje automático de barcos.

- g) Interactividad (I); autonomía (I); adaptabilidad (O): la tradición india habla de la fuerza irresistible del dios Jagannātha, que arrasa con todo lo que encuentra a su paso, acabando incluso con la vida de sus fieles, que, en las procesiones, fallecen bajo las ruedas de los carros que portan las estatuas que lo representan, pues nada puede detenerlo.
- h) Interactividad (I); autonomía (I); adaptabilidad (I): el ser humano es el ejemplo citado por Floridi y Sanders.

La concurrencia de estas tres notas (interactividad, autonomía y adaptabilidad) en un ente enfrentado a una decisión moral nos sitúa, según estos autores, ante un agente moral, y las mismas podrían predicarse de los sistemas de inteligencia artificial más avanzados, en los que faltan, sin embargo, metas, objetivos y cualquier tipo de estado intencional, pues tampoco son libres ni responsables de sus acciones.

Floridi y Sanders advierten de que, si a las citadas características de interactividad, autonomía y adaptabilidad se pretende añadir cualquier otro requisito esencial al concepto de agente moral pleno, dicho requisito ha de estar libre de toda discusión fenomenológica, eliminando así todos los aspectos psicológicos y subjetivistas del concepto. Con ello se «garantiza que el análisis se basa realmente solo en lo que se acepta como observable y no en alguna especulación psicológica». Floridi y Sanders (2004) consideran que este enfoque fenomenológico es una fortaleza de su planteamiento acerca del agente moral.

En esta línea argumental, la concurrencia de estados intencionales (metas, objetivos, deseos, etc.) puede ser una condición agradable para el sujeto, pero es innecesaria para su caracterización como agente moral. Dictaminar sobre la existencia de estados intencionales presupone la disponibilidad de algún tipo de acceso privilegiado a los estados mentales o intencionales del agente, que, aun siendo posible, no se puede garantizar en la práctica en cada caso concreto, hasta el punto de que su existencia acaba por presumirse en las relaciones humanas cotidianas. Lo decisivo es que los agentes realicen acciones morales que «jueguen» al juego moral; que sepan que lo hacen o no será determinante de su responsabilidad, pero es una cuestión que atañe a un momento posterior: el de la determinación y atribución de las consecuencias.

Algo similar ocurre con la objeción de la libertad y, en general, con todas las objeciones basadas en cualquier estado interno especial disfrutado solo por los seres humanos. Para Floridi y Sanders (2004: 14), los sistemas de inteligencia artificial autónomos ya son libres en el sentido de que no se comportan como sistemas deterministas, y eso «nos salva de la horrible perspectiva de tener que entrar en el espinoso debate sobre lo razonable del determinismo, una

infame zona libre, de interminable disputa» y que, por lo tanto, no puede considerarse un requisito esencial del agente moral.

3.4.4. Una moralidad sin mente

En una posición por completo opuesta a la de Floridi y Sanders, se ha sostenido que es esencial que el agente moral se sepa a sí mismo y se muestre a los demás como alguien con una identidad propia, que se conciba a sí mismo (y a los otros) no solo como sujeto, sino también como sujeto racional que actúa intencionalmente y posee sentido de la responsabilidad y de rendición de cuentas (Rodríguez, 2012: 127-138).

Floridi y Sanders insisten, tal y como ya se ha expuesto, en que tales estados cognitivos no añaden nada a la hora de calificar una acción como moral, y postulan separar esos elementos de orden psicológico-cognitivo (*responsibility*) de la cuestión de la rendición de cuentas (*accountability*), atribuyendo esta última a los sistemas de inteligencia artificial en los que concurra el triple requisito (autonomía, interactividad, adaptabilidad) antes señalado, que es el que permite la identificación del agente moral. Se trasladaría así la cuestión de la responsabilidad al campo de la evaluación del acto, pues equiparar identificación y evaluación constituye un «atajo, una falacia jurídica». La responsabilidad es, pues, una consecuencia que permite distinguir un agente moral de una mera causa eficiente, pero no es un requisito definitorio de la esencia de aquel.

Las razones que llevan a esta ampliación del concepto de agente moral para comprender en él a los agentes artificiales son, para estos autores, de orden práctico: a cierto nivel, los requisitos de autonomía, interactividad y adaptabilidad pueden concurrir en entes no humanos. Así ocurre, por ejemplo, en animales de salvamento que se enfrentan a una acción con consecuencias morales, lo que da lugar a reproches (correcciones o castigos) por parte de sus cuidadores, pero no se plantea respecto de ellos una rendición de cuentas, que en todo caso correspondería al ser humano (al «poseedor del animal o al que se sirve de él», según nuestro Código civil). Por el contrario, en el caso de los sistemas de inteligencia artificial, pretender un reproche (corrección o castigo) ante una mala acción tendría el mismo alcance que construir muros más altos en las riberas de un río tras una crecida devastadora, aunque, en sentido contrario, un «castigo» como un apagado definitivo o una reprogramación (como se hace con un sistema de *software* defectuoso) sería inadmisibles al tratarse de un ser vivo. ¿Qué razones hay, pues, para atribuir responsabilidad a los sistemas de inteligencia artificial?

3.4.5. Moralidad distribuida: la complejidad como causa

La concepción tradicional del conocimiento como el producto de una mente individual rompe, a finales del siglo xx, con las aportaciones de Edwin Hutchins acerca de lo que denominó «cognición distribuida». Muchas de las manifestaciones de los procesos cognitivos, a fin de lograr efectividad práctica, se concretan en actuaciones cooperativas y son producto de la interacción colectiva de varios sujetos, ambientes y herramientas, de suerte que

[...] la cognición es inseparable de la interacción con el mundo, y actúa de forma contextualizada y concreta. Partiendo de esto, es posible constatar que la cognición se beneficia de la interacción entre humanos y no-humanos. Es decir, [se] trata de la idea de que la producción de conocimiento no está centrada en un único individuo, sino en la interacción entre los individuos y las herramientas presentes en el ambiente en que estos viven y que auxilian en la producción de conocimientos de forma colaborativa (Regis, Timponi y Maia, 2011).

Desde esta perspectiva, es evidente que la tecnología informática, en general, y, más en concreto, los sistemas de inteligencia artificial se han integrado de manera definitiva en el proceso humano de formación del conocimiento.

La traslación de estas ideas al ámbito de la moralidad y sus relaciones con la robótica y la tecnología en general ha dado lugar al concepto de moralidad distribuida, con el que autores como Lorenzo Magnani y los ya citados Floridi y Sanders se refieren a acciones con resultados relevantes a nivel moral, realizadas por sistemas multiagentes, ya sean de naturaleza exclusivamente humana —como en el caso de las personas jurídicas— o híbrida humano-tecnológica.

Esa idea de la moralidad (y de la cognición) distribuida guarda cierto paralelismo con el problema de la determinación de responsabilidad, que cobra especial relieve en aquellos casos en los que junto a agentes humanos intervienen robots o sistemas de inteligencia artificial con un elevado grado de autonomía y autoplanificación, pues esto puede dificultar en gran medida la determinación del sujeto causante del daño. Aquí, a la intervención del fabricante, del operario y del usuario debe añadirse el papel del programador y, en el caso de robots o sistemas con posibilidad de aprendizaje, el instructor en ese proceso.

A esta cuestión hace referencia expresa el punto 28 de las *Recomendaciones a la Propuesta de Resolución* del Parlamento Europeo, que manifiesta que

[...] una vez que las partes responsables hayan sido identificadas en última instancia, su responsabilidad será proporcional al nivel real de las instrucciones dadas a

los robots y a su autonomía, por lo que cuanto mayor sea la capacidad de aprendizaje o la autonomía, menor será la responsabilidad de las otras partes, y cuanto más larga haya sido la *educación* del robot, mayor será la responsabilidad de su *profesor*.

Hay que tener muy presente, respecto a este último punto, que no se trata de la ejecución repetitiva de instrucciones recibidas en el proceso de *aprendizaje*, pues las competencias adquiridas a través de la *educación* de un robot no deberían confundirse con las estrictamente dependientes de su capacidad de aprender de modo autónomo.

Estas ideas (complejidad, autonomía, aprendizaje, carácter impredecible de las acciones) deben ponerse en relación con la ampliación de los contextos en los que los robots y los sistemas de inteligencia artificial desempeñan sus funciones, pues las barreras físicas de separación que permiten mitigar riesgos en el ámbito de la robótica industrial (campo primigenio de la robótica) son impensables en el caso de los robots asistenciales, médicos, protésicos, o de los vehículos autónomos. Al mismo tiempo, la ampliación de los contextos de utilización de esta tecnología conlleva la ampliación del espectro de los daños que pueden producirse (no solo físicos, sino también morales y patrimoniales) y de las eventuales víctimas (terceros y los propios usuarios).

Claro que, además de esta complejidad contextual, existe otra de orden tecnológico que puede asimismo contribuir a incrementar la dificultad en el deslinde de responsabilidades: en efecto, las técnicas matemáticas empleadas para la formulación de algoritmos de los sistemas de inteligencia artificial provocan lo que se ha denominado «ausencia de trazabilidad de las decisiones» (Barrio y García-Prieto, 2018: 10-12). Como ejemplo, se cita el caso de las redes neuronales artificiales de tipo convolucional empleadas por los sistemas de reconocimiento visual, en las que se asiste a la «emergencia de la independencia (o de agencia)» en la toma de decisiones: durante su aprendizaje, cada neurona recibe múltiples entradas, cada una de las cuales tiene su «peso» específico. Equivalentes funcionales de las fuerzas sinápticas de las neuronas biológicas, esos «pesos» son coeficientes que serán calibrados durante el aprendizaje del algoritmo. Una vez que el aprendizaje haya concluido para reconocer una imagen, una función matemática relaciona aquellos coeficientes con el valor atribuido a determinados píxeles y, aunque es evidente que unos píxeles tienen mayor valor que otros, la determinación de cuáles ha tenido en cuenta el algoritmo resulta matemáticamente imposible.

Y así, concluye García-Prieto (2018: 12), «el algoritmo adquiere cierta distancia frente al diseñador o al programador del mismo [...] y toma sus propias decisiones».

Estas nociones de complejidad externa (contextual) e interna (tecnológica) se traducen, en el Proyecto de informe y en la Resolución del Parlamento de la Unión Europea, en las de mayor autonomía, capacidad de aprendizaje y menor responsabilidad de agentes, como el propietario, el programador o el fabricante, y lleva a considerar al propio robot como un agente más, integrado en un sistema complejo, de suerte que, según el Proyecto:

[...] cuanto más autónomos sean los robots, menos se los podrá considerar simples instrumentos en manos de otros agentes [...] [y] como consecuencia de ello, resulta cada vez más urgente abordar la cuestión fundamental de si los robots deben tener personalidad jurídica.

Por su parte, la Resolución reconduce esa cuestión, no a la atribución de la personalidad jurídica, sino, de un modo más general, a

[...] la cuestión de si la normativa [...] sobre responsabilidad es suficiente o si se requieren normas y principios específicos que aporten claridad sobre la responsabilidad jurídica de los distintos agentes y su responsabilidad por los actos y omisiones de los robots cuya causa no pueda atribuirse a un agente humano concreto.

3.4.6. La cuestión de la responsabilidad jurídica

La complejidad, traducida en una mayor autonomía de los sistemas robóticos y de inteligencia artificial, así como la proliferación de los contextos de uso y la especial trascendencia de los derechos eventualmente afectados, elevan la seguridad y la responsabilidad a categorías fundamentales en los estudios jurídicos sobre estas materias, hasta el punto de haberse acuñado el término «responsibility gap».

La sociedad se enfrenta a un dilema: renunciar al desarrollo de la robótica y de la inteligencia artificial «o enfrentarse a una brecha de responsabilidad que no puede salvarse con los conceptos tradicionales de atribución de responsabilidades» (Matthias, 2004: 1).

Desde luego, una pretensión de prohibición de desarrollo de la industria robótica estaría fuera de la realidad, por lo que la cuestión se traslada a la elección del mecanismo jurídico adecuado para salvar esa brecha en el esquema actual de responsabilidad.

Una primera opción sería el mantenimiento del régimen ya conocido de la responsabilidad objetiva, imponiendo la misma, de ser necesario, con carácter

solidario a cargo de todos los sujetos implicados como contrapartida de las ventajas económicas, tecnológicas y productivas derivadas de la fabricación, entrenamiento, comercialización y utilización.

Si la tecnología alcanza tal grado de desarrollo que la complejidad de funcionamiento se torna un riesgo ante la imposibilidad de determinar quién es el causante del daño, la sociedad, el Derecho, puede exigir responsabilidad a todo aquel que se beneficie de la puesta en circulación de aquel riesgo. Para esta concepción, los eventuales obstáculos derivados del elevado coste de las demandas por daños contra la industria pueden compensarse con un sistema de seguros. Frente a ello se ha afirmado

[...] que los fabricantes del sector difícilmente estarían en capacidad de anticipar y, por lo tanto, de internalizar los costos, potencialmente elevados, de las controversias civiles para el resarcimiento de los daños causados (Palmerini, 2017: 53-97).

Profundizando en estos aspectos económicos, se cita el caso de la grave crisis del sector de fabricación de pequeños aviones, al borde de la desaparición en Estados Unidos por los inasumibles costes de las demandas civiles interpuestas por los usuarios o por sus herederos tras producirse un accidente (Calo, 2011: 132). La General Aviation Revitalization Act («GARA») de 1994 concedió a los fabricantes la inmunidad frente a demandas por defecto de fabricación transcurridos dieciocho años desde la fabricación. El acta supuso la salvación del sector de la aviación comercial de aviones de menos de veinte plazas y, según algunos estudiosos, el resultado, tras más de treinta años de vigencia de la norma, se traduce en un aumento de la tasa de producción y una reducción de la tasa de accidentes.

Un ejemplo puede poner de relieve algunas contradicciones del sistema: una avioneta pilotada por su propietario despegue después de comprobar que no existe ninguna incidencia meteorológica. Tras una hora de vuelo, el piloto recibe el aviso de la proximidad de una tormenta. El piloto toma el mejor rumbo posible de acuerdo con los datos disponibles (el que hubiera tomado cualquier experto en sus mismas circunstancias), pero pese a ello no consigue evitar la tormenta y sufre un accidente mortal. ¿Cambiaría algo de tratarse de una avioneta no tripulada, controlada por un sistema de inteligencia artificial? Si el sistema ha tomado la mejor decisión posible (la que hubiese tomado cualquier experto humano en esas mismas circunstancias), ¿debe haber una responsabilidad objetiva del fabricante? ¿Y del programador o entrenador del sistema?

El de la aviación civil no es el único sector en el que impera el *caveat emptor*, esto es, que le corresponde al comprador desplegar la debida diligencia para examinar el objeto comprado, de modo que la responsabilidad recaiga sobre él. Ryan Calo se refiere también a la ausencia de responsabilidad de los fabricantes de armas por los daños que puedan causar sus usuarios, o de los proveedores de un servicio web por los contenidos difundidos por sus usuarios.

Una segunda orientación defiende un sistema de limitación de responsabilidad del fabricante. A este sistema, según explica Palmerini (2017: 53-97), se le atribuye una doble función: por un lado, la de promover la innovación en la investigación y en la industria robótica, reduciendo el temor ante los elevados costos ligados a la participación en las controversias civiles y, por el otro, la de garantizar la inmunidad a los fabricantes frente a eventos dañinos que no habrían podido ser evitados usando la debida diligencia al diseñar el producto y al informar al consumidor de sus riesgos potenciales.

La exención de responsabilidad del fabricante desplaza el problema a la figura del propietario y/o usuario, y con ello la necesidad de distinguir entre el uso con fines personales y privados, y que tiene una finalidad empresarial o profesional. En el primer caso, el consumidor final

[...] no tendría modo de internalizar los costos ligados a acciones de responsabilidad y debería estar especialmente protegido frente a esta eventualidad, principalmente si se encuentra en condiciones de fragilidad personal (Palmerini: 2017).

Este sería el caso, por ejemplo, de robots de asistencia a enfermos o ancianos o de vigilancia de menores. Esa circunstancia podría ser contraproducente para la propia industria, pues podría desincentivar, en general, su difusión en esos sectores.

Una tercera orientación es la que postula la atribución a los robots y sistemas de inteligencia artificial (que cumplan aquellos requisitos de autonomía antes mencionados) de personalidad jurídica, a la que se la vincularía la titularidad de un fondo patrimonial constituido por aportaciones de aquellos sujetos (fabricantes, programadores, entrenadores, etc.) que se beneficiarían de la exención de responsabilidad. Esta, como ya hemos visto, es una de las propuestas del Proyecto de informe y de la Resolución del Parlamento Europeo (a largo plazo) en el caso de los robots autónomos más complejos, considerándolos, así, como personas electrónicas con derechos y obligaciones específicos, incluida la obligación de reparar los daños que puedan causar, lo que se aplicaría a los supuestos en que los robots pudieran tomar decisiones autónomas inteligentes o interactuar con terceros de forma independiente.

3.4.7. Razones para una negativa

Tras la publicación del *Proyecto de informe con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica* y la Resolución del Parlamento Europeo, se sucedieron diversas reacciones contrarias a la propuesta de atribuir personalidad jurídica a los robots y sistemas de inteligencia artificial.

3.4.7.1. *Las objeciones a su participación en la toma de decisiones con alcance moral*

Desde un punto de vista doctrinal, se ha discutido, no ya la eventual personificación de los robots, sino incluso la posibilidad misma de que sistemas artificiales intervengan en la toma de decisiones con trascendencia ética. Así, se ha propuesto establecer una moratoria a la comercialización de robots que tengan habilidades de razonamiento ético (Van Wynsberghe y Robbins, 2018: 1-20), rechazando que la pretendida complejidad de su funcionamiento y la imprevisibilidad de sus reacciones (más allá de las instrucciones proporcionadas por el programador) sea una causa suficiente para ello. Al contrario, se argumenta que tales artefactos deberían tener vetado el acceso a los contextos en los que pueda haber consecuencias dañosas para los seres humanos a causa de esa complejidad.

3.4.7.2. *La carta abierta de la sociedad civil a los gobernantes de la Unión Europea*

También ha habido reacciones contrarias a aquella posibilidad desde el seno de la sociedad civil. Un grupo de doscientos cincuenta y siete expertos, políticos y empresarios suscribió a principios de 2018 (con la intención declarada de proteger la innovación de la Unión Europea y sus valores, así como la seguridad y la salud de las personas) una carta abierta dirigida a los gobernantes de esta institución en la que se rechaza cualquier vía de personificación de los sistemas robóticos y de inteligencia artificial. En un documento abierto a la suscripción popular, se califica de errónea la afirmación de que la autonomía y el carácter impredecible de los robots haga imposible determinar la responsabilidad por daños. Tal afirmación, en opinión de los firmantes de la carta, obedece a una sobrevaloración de las capacidades reales de los robots y a una comprensión distorsionada del verdadero alcance de lo impredecible y de su facultad de aprendizaje.

De esta forma, la carta rechaza:

- 1) La equiparación de los robots a las personas naturales, ya que entonces el robot tendría derechos humanos, lo que estaría en contradicción con la *Carta de los Derechos Fundamentales* de la Unión Europea y el *Convenio Europeo para la Protección de los Derechos Humanos y de las Libertades Fundamentales*.
- 2) La atribución a los robots del estatus de persona jurídica, pues implicaría la existencia de personas humanas detrás de la persona electrónica para representarla y dirigirla, con lo que la carta abierta de la sociedad civil a los gobernantes de la Unión Europea muestra su rechazo a trasladar a las personas electrónicas el modelo orgánico de representación que tradicionalmente se ha aplicado en las personas jurídicas.

3.4.7.3. *El Estudio del Departamento de Política de Derechos de los Ciudadanos, Libertades Fundamentales y Asuntos Constitucionales del Parlamento Europeo*

La Dirección General de Políticas Interiores de la Unión, en ejercicio de sus funciones informativas al Parlamento Europeo, encomendó un estudio que supuso la primera y radical negativa, desde un ámbito institucional, a la personificación de los sistemas de inteligencia artificial.

Este estudio fue elaborado por Nathalie Nevejans a petición de la Comisión de Asuntos Jurídicos del Parlamento Europeo y se publicó en octubre de 2016. En él se analizan las cuestiones éticas más relevantes planteadas por el *Proyecto de informe con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica*.

Respecto a las materias concretas que afectan a nuestro trabajo, el Estudio se acoge al principio de precaución y afirma que, si los robots llegaran a ser algún día seres conscientes, la humanidad, en su forma actual, estaría condenada a colapsar, rendida ante una forma de inteligencia más poderosa, rápida, perfecta e inmortal, por lo que propone que toda investigación científica encaminada a inculcar la conciencia a un sistema de inteligencia artificial, o que de manera indirecta tuviese ese efecto, sea considerada potencialmente peligrosa para la humanidad y se regule de un modo estricto.

En la misma línea, Nathalie Nevejans (que no en vano encabeza la lista de las personalidades firmantes de la carta abierta antes mencionada) aborda el tema del eventual reconocimiento de cualquier tipo de personalidad jurídica a

los sistemas de inteligencia artificial. La rúbrica del apartado en el que el Estudio se ocupa de esta materia anuncia con claridad la posición de sus autores: «Incongruencia de considerar a los robots como personas jurídicas responsables». A partir de esta proclamación de principios, y desde una perspectiva humanista, el Estudio opta por la negativa a reconocer a los sistemas de inteligencia artificial como sujetos de derecho, calificando tal propuesta de inútil e inapropiada.

El Estudio asimila la atribución de personalidad jurídica a una pretensión de humanización, vinculando aquella a la existencia de cierta forma de autoconciencia y a la capacidad de sentir (en la línea de las posiciones utilitaristas a las que se ha hecho referencia en este trabajo). Al mismo tiempo, reconoce que el Proyecto normativo de la Unión Europea no vincula, en realidad, la atribución de la personalidad (al menos no en el momento actual) a la aparición de una autoconciencia artificial, sino a un simple objetivo operativo: hacer a los sistemas de inteligencia artificial responsables de sus actos a nivel jurídico.

Nathalie Nevejans rechaza la posible asimilación de la pretendida personalidad electrónica a la figura de las personas jurídicas, argumentando que, aunque, de hecho, tampoco concurre en tales entes la condición humana (que los convertiría de forma natural e inmediata en personas —naturales—), respecto a ellos se ha acudido históricamente a la ficción jurídica de considerarlos sujetos de derechos porque «en última instancia, se trata [...] de una persona física que insufla vida jurídica a una persona jurídica y sin la cual, esta última es una mera cáscara vacía» (Nevejans, 2016: 14).

Tras analizar el posible paralelismo entre una eventual persona electrónica y una persona jurídica, Nathalie Nevejans plantea dos posibles escenarios: o bien existe una persona física que es el verdadero actor legal detrás del robot, o bien el propio robot es un actor legal.

En el primer caso, argumenta el Estudio, si consideramos que hay una persona detrás del robot autónomo, entonces esta representaría a la persona electrónica, la cual, desde el punto de vista jurídico, sería una construcción intelectual ficticia. Nevejans califica de incongruencia construir un mecanismo tan sofisticado para llegar a lo que ella considera un resultado inútil.

La segunda opción está más en la línea de la Propuesta de Resolución, que parece proclive a eliminar por completo la presencia humana. Y aquí, Nathalie Nevejans reitera el mismo argumento y se pregunta: «¿cómo puede una mera máquina, un cadáver desprovisto de conciencia, de sentimientos, o de voluntad propia, convertirse en un actor legal autónomo?» (Nevejans, 2016: 14). Concluye diciendo que es imposible (desde un punto de vista científico, jurídico, e incluso ético) que en los próximos quince años (plazo al que se refiere el

apartado 25 de la Propuesta) «un robot [pueda] participar en la vida legal sin que un ser humano mueva los hilos».

El Estudio califica de absurdo, e incluso de potencialmente peligroso, para el mercado emergente de robots la atribución de personalidad jurídica a los sistemas de inteligencia artificial. La línea argumental de Nathalie Nevejans parte de la afirmación de la Propuesta según la cual la atribución de personalidad jurídica a los robots significaría que sus derechos y deberes deberían ser respetados. A continuación se plantea qué derechos se le otorgarían a un robot, cuestión a la que responde con una enumeración de derechos tales como el derecho a la vida (es decir, a la no destrucción), a la dignidad, a la igualdad con la humanidad, a la educación, a la salud, a jubilarse y a recibir una remuneración.

Más adelante, Natalie Nevejans pone como ejemplo dos tipos de robots que, en algunos casos, ya pueden ser considerados como autónomos inteligentes. Se pregunta entonces si asignar derechos significa que un robot desplegado en un entorno hostil podría negarse a cruzar una zona peligrosa, o que un robot para el cuidado de la salud exigiese el pago de un salario antes de empezar a cuidar a un anciano. Si la respuesta es afirmativa, concluye el Estudio, ¿quién estaría interesado en comprar robots que pudiesen negarse a funcionar?

Denegada la posibilidad de personificación jurídica de los sistemas de inteligencia artificial, el Estudio recuerda que si el problema principal que se trata de resolver con la asignación de la personalidad es el de la responsabilidad en caso de daños, hay otros sistemas que podrían ser más eficaces a la hora de indemnizar a las víctimas, como el de seguro obligatorio a cargo del fabricante, combinado con un fondo de compensación, tal y como prevé la propia Propuesta de informe (apartado 31).

Para concluir su crítica a la Propuesta, Natalie Nevejans afirma que los defensores de la opción de la personalidad jurídica tienen una visión fantasiosa del robot, inspirada en las novelas de ciencia ficción y el cine, y considera que sería inapropiado y que estaría fuera de lugar crear dicha personalidad jurídica, pues

[...] con ello se corre el riesgo, no sólo de asignar derechos y obligaciones a lo que no es más que una herramienta, sino también de derribar las fronteras entre el hombre y la máquina, desdibujando las líneas entre lo vivo y lo inerte, lo humano y lo inhumano (Nevejans, 2016: 14).

Todo ello, concluye, acabaría por degradar al ser humano a la categoría de máquina y por poner en tela de juicio los fundamentos humanistas de Europa.

3.4.7.4. El Dictamen del Comité Económico y Social Europeo

El 31 de agosto de 2017, el *Diario Oficial* de la Unión Europea publicó el *Dictamen del Comité Económico y Social Europeo sobre la Inteligencia artificial: las consecuencias de la inteligencia artificial para el mercado único (digital), la producción, el consumo, el empleo y la sociedad*. En él, el CESE hace un recorrido por las materias en las que mayor repercusión puede tener el desarrollo e implantación de la inteligencia artificial. Para ello, parte de la afirmación de que la cuestión ha de analizarse no solo desde el punto de vista técnico, sino también desde la perspectiva de la ética, la sociedad y la seguridad, y, asimismo, reconoce la importancia económica del sector con un volumen de mercado estimado en torno a los 664 millones de dólares estadounidenses y que se espera que aumente a 38 800 millones en 2025. Cabe destacar que estas cifras, tomadas directamente del citado documento, no parecen correctas, y, así, el artículo «In Focus: How AI is reshaping the future and our investments», disponible en la web de JP Morgan, habla de una cifra, para el año 2019, próxima a los 35 800 millones de dólares, un incremento superior al 44 % respecto del año 2018, y se proyecta que, en el año 2022, el gasto en sistemas de inteligencia artificial sea más del doble, hasta llegar a casi 80 000 millones de dólares, con un crecimiento anual compuesto del 38 % entre 2018 y 2022.

Ética, seguridad, privacidad, transparencia y rendición de cuentas; trabajo, educación y desarrollo de capacidades; (des)igualdad e inclusión; legislación y reglamentación; gobernanza y democracia; guerra; superinteligencia: son las áreas donde el Dictamen considera que la inteligencia artificial puede cobrar mayor relevancia. En el ámbito concreto que afecta a esta tesina, opta igualmente por una llamada a la prudencia y, de forma expresa, se hace eco de la postura defendida por Stephen Hawking, quien opinó que el desarrollo de una inteligencia artificial fuerte significaría el fin de la humanidad, pues evolucionaría a un ritmo que los humanos no podrían seguir. De ahí que algunos expertos aboguen por lo que llaman un *kill-switch* o *reset-button*, un botón para desactivar o reiniciar los sistemas de inteligencia artificial desbocados o superinteligentes.

Al mismo tiempo, el Dictamen defiende como un principio esencial el control humano (*human-in-command*), con un marco de condiciones que regule el desarrollo y uso responsable y seguro de la inteligencia artificial, de manera que las máquinas continúen siendo máquinas, y los humanos conserven en todo momento el dominio sobre ellas.

Como corolario de tal línea argumental, el Dictamen se pronuncia en contra de la introducción de cualquier tipo de personalidad jurídica para los robots o los sistemas de inteligencia artificial, argumentando que socavaría los

efectos correctores preventivos de la legislación en materia de responsabilidad, generaría un riesgo moral tanto en el desarrollo como en la utilización de la inteligencia artificial y daría lugar a un posible uso indebido.

3.4.8. Los contraargumentos

3.4.8.1. *La confusión con el ser humano*

Las posturas contrarias a la personificación de los sistemas de inteligencia artificial suelen partir de una perspectiva antropológica de la persona que la equipara exclusivamente con el ser humano, al igual que hacen los modelos que antes (al combinar los elementos antropológico, biológico y filosófico/psicológico) caracterizamos como «Persona = ADN humano (1), racional/consciente/sintiente (o) y ente con función jurídica (1)» y como «Persona = ADN humano (1), racional/consciente/sintiente (o) y ente con función jurídica (o)».

Desde esa perspectiva, resulta lógico calificar —como hace el Estudio— de absurda la pretensión de atribuir a los robots derechos tales como el derecho a la vida, a la dignidad o a la igualdad con la humanidad. Ocurre, sin embargo, que esta argumentación parece descansar en un sofisma de falsa premisa, pues en ningún punto del Proyecto de informe se propone la atribución de tales derechos (vida, dignidad, igualdad) a los robots, sino que de manera expresa se habla de «la creación de una nueva categoría, con sus propias características y repercusiones en lo que se refiere a atribución de derechos y obligaciones, incluida la responsabilidad por daño», sin que en ningún momento se proponga esa equiparación con derechos que, por su naturaleza, forman parte del núcleo esencial de los derechos humanos (véase, sobre esta cuestión, el punto 5.3.3).

3.4.8.2. *La confusión con las personas jurídicas. Referencia al esquema societario como posible cobertura para una personalidad jurídica independiente de los sistemas de inteligencia artificial*

Las posturas negativas antes expuestas rechazan también la equiparación de los sistemas de inteligencia artificial con las personas jurídicas. Así, la carta abierta dirigida a la Unión Europea argumenta que toda persona jurídica implica la existencia de personas humanas para representarla y dirigirla, y Nathalie Nevejans, en el citado Estudio, afirma que considerar a las corporaciones, sociedades e instituciones como sujetos de derechos es una ficción jurídica que no oculta

que, en última instancia, siempre existe una persona física que insufla vida jurídica a la persona jurídica, sin la cual esta última sería «una mera cáscara vacía».

Esta crítica, sin embargo, fracasa, por un lado, porque confunde a la persona jurídica con su órgano de representación y, por otro, porque la configura como una mera ficción, ignorando que, al atribuir personalidad jurídica a instituciones, asociaciones y corporaciones, el legislador no interviene aquí brutalmente por motivos de oportunidad, sino que no hace más que secundar en términos jurídicos lo que ya existe en la concepción práctica social. Del mismo modo que «el hombre no es sujeto por construcción de la ley, así tampoco la colectividad es producto de la técnica jurídica» (Ferrara, 2006: 260).

En la misma dirección, el argumento de que siempre ha de existir una persona física que «insufle vida» a la persona jurídica ha sido puesto en entredicho por algunos autores que tratan de encontrar acomodo para una personalidad jurídica independiente de los sistemas de inteligencia artificial dentro de los esquemas legales vigentes en la actualidad para las personas jurídicas mercantiles.

Así, Shawn Bayern estudia la posibilidad de utilizar la cobertura de algunas figuras societarias como contenedor para una personalidad jurídica de sistemas robóticos o de inteligencia artificial sin respaldo humano (Bayern, 2015: 1-20). La regulación estadounidense de las LLC (*Limited Liability Company*, un esquema societario a medio camino entre las sociedades de tipo personalista y las de capital del Derecho europeo continental), recogida en la Uniform Limited Liability Company Act, permite, según explica Bayern, constituir una LLC con un solo socio, atribuir la administración a un sistema de inteligencia artificial y, posteriormente, desvincular al socio de la sociedad. Esta subsistiría después durante un plazo de noventa días, salvo disposición contraria en los estatutos sociales, permitiendo así establecer por vía estatutaria la continuación de la sociedad como sujeto de derecho sin ningún respaldo humano (que le «insufle vida») y gobernada por un sistema artificial.

Trabajos posteriores (Bayern *et al.*, 2017: 135-162) estudian esa misma posibilidad en diversos ordenamientos jurídicos europeos. Así, encontramos que, en el Derecho alemán, la regulación de la *Gesellschaft mit beschränkter Haftung* (en abreviatura, GmbH, figura similar a nuestra sociedad de responsabilidad limitada) permite (al menos teóricamente y por las mismas vías que después veremos para el Derecho español) la existencia de «sociedades sin socios» (la cuestión ya había sido planteada por Max Hachenburg en un trabajo de 1915). No obstante, exige siempre la presencia de al menos una persona física al frente del órgano de administración, de modo que no serviría para albergar la subjetividad jurídica independiente de un sistema de inteligencia artificial. Algo similar ocurre en el Derecho suizo (donde se ha estudiado también la

utilización de fundaciones como medio de personificación jurídica de sistemas de inteligencia artificial) y en el resto de los ordenamientos de corte continental. El Derecho inglés, más flexible, con la regulación de las *Limited liability partnerships* (LLP), permite la atribución de la gestión a un sistema de inteligencia artificial, si bien exige que tenga un mínimo de dos socios, e impone responsabilidad ilimitada por las deudas sociales a cualquier socio que consienta que la sociedad continúe con sus operaciones durante más de seis meses con menos de dos socios. Ahora bien, la ley inglesa no impone la disolución por esa causa, lo que da pie a la continuidad de la sociedad, que podría, aunque fuera de manera transitoria y en situación jurídica de pendencia, actuar en el tráfico jurídico, incluso desvinculada de toda responsabilidad patrimonial subsidiaria de los socios, por ejemplo, en caso de fallecimiento simultáneo o en un período inferior a seis meses de los dos únicos socios.

El estudio de esta cuestión permanece inédito en el Derecho español. Refiriendo esa cuestión al esquema de las sociedades de capital (anónimas o de responsabilidad limitada), se deben analizar dos planos para determinar si es posible utilizar tal esquema como cobertura para dotar de subjetividad jurídica a los sistemas de inteligencia artificial:

- El plano de la titularidad del capital social: desde hace varias décadas, nuestro Derecho admite y regula las sociedades (anónimas y limitadas) unipersonales. Los supuestos de desaparición del socio único se contemplan como situaciones transitorias (piénsese, por ejemplo, en el tiempo transcurrido desde el fallecimiento del socio único hasta la aceptación de la herencia por sus herederos), pero existen algunos supuestos aparentemente paradójicos que podrían servir de soporte a una personalidad jurídica independiente de respaldo humano para sistemas de inteligencia artificial. Así, por ejemplo, es posible que una persona física constituya como único socio una sociedad limitada, aportando al capital social (como aportación no dineraria) un sistema de inteligencia artificial de ventas *online* que acepta el pago en línea y que utiliza los fondos recaudados para reponer *stocks* y gestionar los envíos de los objetos vendidos. Una vez cumplidos todos los requisitos legales para la constitución de la sociedad, son varias las vías para llegar a aquellas situaciones paradójicas de «sociedad sin socio»: por ejemplo, mediante cesión de todas las participaciones de la sociedad a otra sociedad cuya única socia es la primera o mediante sucesión *mortis causa* a la que concurre la propia sociedad como heredera, etc. Podrá alegarse que tales casos pueden reconducirse al artículo 141 del *Texto refundido de la Ley de Sociedades de capital*, que impone un plazo de tres años para que cualquier

interesado pueda solicitar la amortización de esas participaciones, lo que acabaría por provocar la disolución de la sociedad al constituir tales participaciones todo el capital social. No obstante, debe destacarse que dicho precepto ni impone una disolución *ex lege* ni opera de manera automática, y solo lo hace una vez transcurridos tres años.

- El plano de la gestión de la sociedad: en este segundo aspecto, las dificultades parecen ser mayores, pues cualquiera que sea la forma que se elija para la administración de la sociedad (administrador único, solidarios, mancomunados o consejo de administración), necesariamente recaerá sobre personas físicas o jurídicas (y estas últimas deben designar a una persona natural para el ejercicio permanente de las funciones propias del cargo). *A priori*, no es, pues, posible a nivel legal, excluir la presencia humana del órgano de administración. Excede del objetivo de este trabajo el estudio de las cláusulas estatutarias por las que podría imponerse que el órgano de administración de la sociedad tuviera la obligación de respetar o no interferir en la ejecución de actos de gestión por un algoritmo o sistema de inteligencia artificial, pero ha de destacarse que incluso aunque fuese materialmente posible que un algoritmo ejecutase todos los actos de desarrollo del objeto social (por ejemplo, compra-venta *online*, *trading* algorítmico), e incluso los de trámite o gestión (obligaciones fiscales, administrativas, etc.), siempre recaería sobre los administradores la eventual responsabilidad (artículos 225 y siguientes de la Ley de sociedades de capital) derivada de aquellos actos frente a la sociedad y acreedores (no mencionamos a los socios en este punto, dando por supuesto que estos no existirían según lo expuesto en el apartado anterior). Sí cabe, sin embargo, plantearse el caso en que, constituida la sociedad con un administrador único y privada de socios en la forma expuesta, quede en situación de acefalia por la dimisión del administrador, etc., y aquí tendríamos no una mera cáscara vacía (como en tono poético pretendía criticar Nathalie Nevejans), sino un verdadero sujeto de derechos (una sociedad, con el registro mercantil cerrado y sin posibilidad de presentar cuentas, sí), desarrollando, *sine die*, su objeto social de modo automatizado sin una persona humana que le «insufle vida».

3.4.8.3. Juridificación de la ciencia ficción o regulación previsora

Otra línea argumental, plasmada en el estudio dirigido por Nathalie Nevejans y contraria a la personificación de los sistemas de inteligencia artificial, hace

hincapié en que los defensores de la opción de la personalidad jurídica tienen una visión fantasiosa de los sistemas de inteligencia artificial, inspirada en el cine y la literatura pero alejada de la realidad científica.

A esta cuestión se refiere, desde una perspectiva más general, en relación con el papel de la bioética en el campo de la biotecnología, Manuel Jesús López Baroni, al tratar lo que califica como «juridificación de la ciencia ficción». Aunque, ciertamente, «se corre el riesgo de regular predicciones fantásticas y de deslizarse sin remisión por la pendiente del ridículo» (López, 2018: 8), una espera contemplativa a la política de hechos consumados podría traducirse tanto en obstáculos al progreso científico y económico como en dificultades para obtener indemnización por daños en caso de accidente, así como en la potencial proliferación, ante la paralización del legislador, de «monstruos» jurídicos (como los estudiados antes) en la búsqueda de cobertura personificada para sistemas de inteligencia artificial.

La dificultad de determinar el momento adecuado para actuar debe ponerse bajo el foco de la realidad. En este sentido, la *Resolución del Parlamento Europeo de 16 de febrero de 2017 con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica* remite estas cuestiones a un momento que sitúa en «el largo plazo». Nathalie Nevejans, en el Estudio citado, señala que eso no sucederá antes de quince años, y en el Proyecto de informe del 31 de mayo de 2016, la Comisión de Asuntos Jurídicos del Parlamento Europeo hablaba de un plazo de diez o quince años, lo cual no parece un horizonte ni muy lejano ni plagado de fantasías, máxime si tenemos en cuenta, por citar un ejemplo, que el coche autónomo más avanzado, en la actualidad en fase de prueba, alcanza el nivel 4 (de un máximo de 5 en la clasificación de la SAE International, Asociación internacional de ingenieros de automoción). Esto supone prescindir tanto de conductor como de usuario entrenado para su manejo, pues el vehículo es autónomo y está preparado para detenerse con seguridad en caso de fallo del sistema principal.

3.4.8.4. Futuro y singularidad

Antes se ha sostenido que el modelo defendido por los detractores de la personificación jurídica de los sistemas de inteligencia artificial encaja en los esquemas «Persona = ADN humano (1), racional/consciente/sintiente (o) y ente con función jurídica (1)» o «Persona = ADN humano (1), racional/consciente/sintiente (o) y ente con función jurídica (o)». Recordemos que en estos esquemas la nota esencial identificativa de la persona es la pertenencia a la especie huma-

na, y son totalmente irrelevantes las nociones de autoconciencia, racionalidad o sensibilidad. Sin embargo, en ocasiones, aquellas tesis acuden a postulados próximos al utilitarismo para negar tal atribución de personalidad.

Así, Nathalie Nevejans argumenta que es fundamental discutir el asunto de la autoconciencia de los robots, y vincula la pretensión de personificación con un intento de humanización similar a la mantenida por los defensores de los animales cuando defienden «que debería asignarse a los animales personalidad jurídica, ya que algunos son seres conscientes y capaces de sufrir y con sentimientos que los separan de las cosas» (Nevejans, 2016: 14).

Así, el *Estudio del Departamento de Política de Derechos de los Ciudadanos, Libertades Fundamentales y Asuntos Constitucionales* del Parlamento Europeo se aproximaría a las posiciones que configuran la persona bajo el esquema «Persona = ADN humano (o), ente racional/consciente/sintiente (I) y ente con función jurídica (o)».

En rigor, debe precisarse que el Proyecto de informe y la Resolución del Parlamento Europeo se refieren al supuesto hipotético de una eventual autoconciencia de la inteligencia artificial de una forma incidental. Este es el texto del Proyecto de informe:

Considerando que, hasta que los robots sean conscientes de su propia existencia o sean fabricados con esa cualidad, si es que ese momento llega algún día, debe entenderse que las leyes de Asimov van dirigidas a los diseñadores, fabricantes y operadores de robots, dado que dichas leyes no pueden traducirse en código de máquina.

Y la Resolución se pronuncia sobre esta materia en términos más moderados y de forma incluso más superficial:

Considerando que existe la posibilidad de que a largo plazo la inteligencia artificial llegue a superar la capacidad intelectual humana.

Resulta, pues, que cuando el Proyecto de informe (y la posterior Resolución del Parlamento Europeo) propone la atribución de personalidad jurídica electrónica a los robots autónomos, no lo hace en atención a su (eventual e hipotética) autoconciencia, sino independientemente de ella y para el caso en que lleguen a tomar decisiones autónomas. No es cierto, por lo tanto, que, como pretende Natalie Nevejans, exista un intento de humanización en aquella personificación, ni que esta implique la necesidad de reconocer derechos básicos a los robots. Tales derechos básicos, caracterizados como derechos hu-

manos o fundamentales, son patrimonio exclusivo del ser humano, ya se acuda para fundamentarlos a la dignidad humana desde posiciones iusnaturalistas, o a criterios objetivos desde posiciones historicistas o consensualistas. Por ello, no resulta lícito criticar la pretensión de personificación de los sistemas de inteligencia artificial aduciendo que implica un peligro de humanización de las máquinas.

La cuestión de la (eventual) autoconciencia de los sistemas de inteligencia artificial se relaciona con la aparición de la llamada «inteligencia artificial fuerte»: un hipotético sistema capaz de emular el total funcionamiento de la mente humana, incluyendo no solo la capacidad de resolución de tareas, sino también los sentimientos y la autoconciencia. Tal concepción es en extremo polémica: por un lado, porque son muchos los escépticos que consideran que ese acontecimiento (calificado como singularidad) no ocurrirá y, desde luego, no tendrá lugar en un futuro inmediato y, por otro, porque entre los defensores de tal posibilidad abundan los que la consideran perjudicial para la humanidad. Ciertamente, las preocupaciones por una inteligencia artificial superinteligente (Russel: 2017, 179-180) no son algo nuevo, y ya el mismo Turing en una comunicación por radio del año 1951

[...] sintió la necesidad de apuntar la siguiente posibilidad: «Si una máquina puede pensar, es posible que piense de manera más inteligente que nosotros, y entonces ¿dónde nos coloca eso? Incluso si pudiéramos relegar a las máquinas a una posición de servidumbre, por ejemplo desenchufándolas en momentos estratégicos, deberíamos, como especie, extraer una lección de humildad []. Este nuevo peligro [] es, sin duda, algo que puede causarnos preocupación».

3.4.8.5. *Tras la barrera humanista: el dataísmo*

La ruptura de los valores del humanismo es uno de los mayores riesgos inherentes a la personificación jurídica de los sistemas de inteligencia artificial, según los detractores de tal propuesta. Así, Nathalie Nevejans (2016) habla de los peligros de derribar las fronteras entre el ser humano y las máquinas, y afirma que el reconocimiento de lo que el Proyecto de informe del 31 de mayo de 2016 (y la Resolución posterior del Parlamento de la Unión Europea) perfila como personalidad electrónica «pondría en tela de juicio los fundamentos humanistas de Europa».

Remedio contra el sometimiento a un plan cósmico y la dictadura de un destino impuesto por la divinidad, el humanismo entregó al ser humano las

riendas de su existencia, sustituyendo, como almacén estructural de la sociedad, la fe en los dioses que había guiado a la civilización durante los siglos anteriores por la fe en la humanidad —sus valores, voluntad, sentimientos, experiencias, inteligencia—, y se concretó en diversas teorías de orden socialista, fascista, de tipo evolutivo y, en el ámbito de la Unión Europea, de orden liberal cristiano.

Las referencias a esos valores humanistas son una constante citada con reiteración a modo de infranqueable barrera en todos los textos contrarios a la consideración de los sistemas de inteligencia artificial como sujetos de derecho. El temor jurídico a esa personificación de la inteligencia artificial, paralelo a la llamada al principio de prudencia en la investigación científica en estos sectores, se debe a que está en juego el principio esencial, fundador, del almacén humanista: la voluntad humana libre. Así (Harari, 2016: 400-431), para la concepción humanista, las tecnologías encaminadas a sustituirla, controlarla, rediseñarla o incluso mejorarla estarían atentando contra la piedra de toque de la civilización.

Ocurre que, tras aquella barrera levantada por el humanismo, se alza hoy una propuesta, el dataísmo, para la cual lo esencial no es un plan divino ni tampoco la voluntad humana, sino que «el dataísmo sostiene que el universo consiste en flujo de datos y que el valor de cualquier fenómeno o entidad está determinado por su contribución al procesamiento de datos» (Harari, 2016: 400).

Las leyes matemáticas que rigen los algoritmos electrónicos son las mismas que regulan los algoritmos bioquímicos, lo que difumina la diferencia entre lo natural y lo artificial, y provoca que «la barrera entre máquinas y animales se desplome» (Harari, 2016: 400). Un sistema de inteligencia artificial determina cuándo, qué, cuántos y a qué precio comprar determinados valores en una negociación *online*. Al otro lado de la pantalla, un bróker determina cuándo, qué, cuántos y a qué precio vender ciertos valores. En tal escenario, donde el humanismo contemplaba una máquina confrontada con la voluntad humana desarrollando sus fines e intereses, el dataísmo describirá dos algoritmos, uno artificial y electrónico, y otro natural y bioquímico, interactuando, contribuyendo al aumento del caudal del flujo de datos.

Y si, para el humanismo jurídico, la voluntad humana (sus intereses) es el núcleo esencial de la subjetividad jurídica, el dataísmo atribuirá tal papel a los algoritmos que contribuyan a aquel flujo universal de información. En el ámbito jurídico, ese flujo se concreta en los datos de los intervinientes en los contratos o, en general, en actos con trascendencia jurídica, de las circunstancias antecedentes y coetáneas del mercado, de las prestaciones (bienes o servi-

cios) intercambiadas, de su valor, de su forma de pago y de la fecha, lugar y normas jurídicas de aplicación. Ese flujo de datos jurídicos se retroalimenta sin pausa desde centros subjetivos de base algorítmica para la imputación de consecuencias jurídicas que, a modo de atractores, al ejecutar activa o pasivamente prestaciones, canalizan información y emiten y reciben datos, siendo indiferente para la regulación jurídica de ese flujo que la base de aquellos algoritmos sea electrónica o bioquímica.

En la actualidad, una concepción dataísta de la personalidad jurídica respondería a un esquema que prescindiría tanto del requisito del ADN humano como de la exigencia de conciencia o inteligencia siguiendo el modelo «Persona = ADN humano (o), racional/consciente/sintiente (o) y ente con función jurídica (o)». Ni la pertenencia a la especie humana ni la existencia de un elemento de raciocinio, autoconciencia o sensibilidad; lo relevante es la contribución al flujo jurídico de datos que realizan esos centros de imputación concebidos, desde esta perspectiva, como personalidades algorítmicas.

Situemos ahora esta referencia al dataísmo junto con el recordatorio de que la razón por la que se ha justificado la conveniencia de atribuir personalidad jurídica a los sistemas de inteligencia artificial deriva de los problemas que plantea la determinación de la responsabilidad ante daños cometidos por tales sistemas a causa de su complejidad y autonomía.

Ante tal planteamiento —un problema de determinación de responsabilidad—, cualquiera podría ver una lección histórica en el hecho de que fue también un problema de determinación de la responsabilidad el que provocó la consolidación de la doctrina canónica de la persona ficta, elaboración doctrinal que pretendía evitar que la pena de excomunión a las ciudades o corporaciones que se rebelaban contra el papa o el emperador recayera sobre cada uno de sus miembros, fuesen culpables o no.

Eso aconteció a mediados del siglo XIII, tan solo unas décadas antes de que el nuevo espíritu del Renacimiento acabase por establecer el humanismo como la argamasa de la cohesión social durante los siglos venideros, lo que arrojaría por el sumidero de la historia la anterior concepción secular de que todo lo existente y por venir era fruto de un plan divino universal.

Entonces, se trataba de determinar la responsabilidad por el daño máximo que se podía ocasionar bajo aquel esquema mental y de organización de la sociedad: el daño a Dios.

Hoy, cuando se plantea la atribución de personalidad a los robots y sistemas de inteligencia artificial, se trata de determinar la responsabilidad por el daño máximo que se puede causar bajo nuestro actual esquema mental y de organización de la sociedad: el daño al ser humano.

4.1. CUÁNDO

La Resolución del Parlamento Europeo sitúa a largo plazo el momento en que se cree una personalidad jurídica específica electrónica que se extienda a aquellos robots autónomos más complejos y, posiblemente, a los que tomen decisiones autónomas inteligentes o interactúen con terceros de forma independiente. Sin duda, esa referencia al largo plazo en la que insiste la Resolución es una respuesta a las llamadas al principio de precaución realizadas por la doctrina y la opinión pública ante las inquietudes generadas por la publicación previa del Proyecto de informe, pero, más que aplazar la cuestión, lo que se hace es remitirla al criterio científico.

Será la ciencia la que dictaminará cuándo esa complejidad, esa autonomía del sistema permitirá su actuación independiente, de modo que sea imposible atribuirle técnicamente a un ser humano el resultado de esa actuación. Sin embargo, para evitar que la personalidad se limite a un contenido tan solo negativo, sería conveniente exigir, además, una mínima capacidad relacional y de autogestión que permita al sistema realizar *online*, al menos, los trámites necesarios para su constitución, el pago de tasas e impuestos, los contratos de suministros y los trámites con las administraciones públicas.

Hasta entonces, es previsible la aparición de soluciones jurídicas exóticas y paradójicas, resultado de sustituir la presencia humana por sistemas de inteligencia artificial en los órganos de capital y de gestión de sociedades mercantiles u otras personas jurídicas clásicas (*cf. supra* 4.4.8.2.).

En cualquier caso, y a falta de acuerdos internacionales, el aspecto temporal puede perder toda relevancia si fuera del ámbito de la Unión Europea otros ordenamientos regulan estas materias, máxime cuando el principio de precaución no actúa de igual forma en el ámbito europeo que en el asiático o el estadounidense.

4.2. DÓNDE

Ante la falta, tal y como se acaba de mencionar, de acuerdos internacionales y respecto a una eventual regulación de la Unión Europea, Thomas Burri (2018) recuerda que queda fuera de la competencia de la Unión Europea imponer una regulación sobre esta materia, pues la determinación, tanto del concepto legal de persona natural como de los requisitos, es competencia de los Estados miembros para que ciertas entidades, sociedades, fundaciones, asociaciones, etc. tengan la condición de sujetos de derecho.

4.3. CÓMO

4.3.1. Los derechos afectados y los principios bioéticos

La Resolución del Parlamento Europeo, en el apartado titulado «Principios éticos», sitúa la necesidad de regulación en el punto de equilibrio entre las ventajas sociales y económicas de la robótica y los posibles riesgos que su implantación y generalización pueden suponer para «la seguridad y la salud humanas; la libertad, la intimidad, la integridad y la dignidad; la autodeterminación y la no discriminación, y la protección de los datos personales». También consagra, de forma expresa, el marco de orientaciones éticas en el que debe basarse no solo la eventual regulación sobre la personalidad jurídica electrónica, sino, en general, la normativa sobre robótica e inteligencia artificial, y lo sitúa en torno a cuatro principios de honda raigambre bioética: los principios de no maleficencia, beneficencia, autonomía y justicia.

Estos principios vertebran también el Borrador de directrices éticas para inteligencia artificial confiable, publicado en diciembre de 2018 a iniciativa de la Comisión Europea. En este documento, el grupo de expertos de alto nivel sobre inteligencia artificial articula sobre aquellos principios las líneas éticas que están llamadas a inspirar la futura normativa europea en esta materia. Para ello, y con el fin de lograr que la implantación de sistemas de inteligencia artificial maximice los beneficios al tiempo que minimice los riesgos, parte del concepto de inteligencia artificial confiable, pues «los seres humanos solo serán capaces de cosechar plena y confiadamente los beneficios de la IA si pueden confiar en la tecnología».

Por inteligencia artificial confiable se entiende aquella que cumple el doble requisito de respetar los derechos fundamentales con un propósito ético y ser técnicamente robusta, tratando de excluir así que la falta de dominio tecnológico pueda causar daños involuntarios.

4.3.1.1. *No maleficencia. No dañar*

Al comentar este principio, el citado Borrador de directrices éticas afirma que, como mínimo, los sistemas de inteligencia artificial deben diseñarse de manera que ni agraven daños existentes ni creen otros nuevos para los individuos, daños que pueden ser físicos, psicológicos, financieros o sociales, de ámbito tanto individual (como los referidos a la libertad, intimidad, etc.) como colectivo (laborales, ambientales, etc.).

Desde aquella óptica individual y concebido como un mecanismo preventivo del daño, el principio de no maleficencia nos sitúa de forma inmediata ante la primera de las leyes de Asimov, pero la afirmación que encierra («Un robot no hará *daño* a un ser humano ni permitirá, por inacción, que un ser humano sufra daño») tiene un soporte de mayor fortaleza: el principio de dignidad humana al que Natalie Nevejans se refiere de forma expresa en el citado estudio y que no en vano aparece recogido en la *Carta de los Derechos Fundamentales* de la Unión Europea del 7 de diciembre de 2000. Ahí se establece que la dignidad humana es el fundamento de todos los derechos y que es inviolable, tal y como resulta también de la *Declaración Universal de los Derechos Humanos* y acaba por plasmarse en las Constituciones de los ordenamientos jurídicos modernos, que la sitúan (así, el artículo 10 de la Constitución española de 1978) como uno de los fundamentos del orden político y de la paz social. Es por ello acertado considerar, tal y como hace el referido estudio, la dignidad humana

[...] como un concepto fundador que podría servir de base jurídica para la robótica [...] [y situarlo] como la fuente de todos los principios que protegen a las personas del daño, permitiendo que los seres humanos estén protegidos contra los daños infligidos a su persona física por los robots.

De esta forma, la dignidad humana está llamada a desempeñar, en el ámbito de la robótica, el mismo papel tuitivo que juega en el de la biotecnología en general y, sin duda, será objeto de referencias expresas, en principios y articulados, de los convenios internacionales que se celebren en esta materia, tal y como ocurre, por ejemplo, respecto a la biología y la medicina en el Convenio de Oviedo de 1997. Esto en el plano jurídico. En el ético, la dignidad humana actúa como patrón de la legitimidad moral de los avances biotecnológicos y limitador de su alcance (Hervás 2001: 97).

Ocurre, sin embargo, que esa relevancia de la dignidad como principio fundador no va acompañada de la concreción de su concepto, por lo que se

corre el riesgo de que se presente como un elemento vacío de contenido; en debates, un arma arrojadiza más próxima a un «protector de tabúes» (Hervás 2001: 98) que a una frontera benefactora.

Perfilando su concepto a partir de su recepción en el Derecho positivo, encontramos (Oheling 2011: 140, 141) que la «desconfianza respecto a las teorías del positivismo mecanicista típico de entreguerras» llevó, en el proceso constituyente alemán, a la normativización de principios prepositivos, reconociendo la dignidad humana como fundamento de la comunidad humana y como límite inviolable frente a la injerencia del poder público, idea que más tarde se generalizará en las Constituciones europeas.

En su desarrollo doctrinal y jurisprudencial, la dignidad se relaciona de forma indisoluble con la autonomía individual, el libre desarrollo de la personalidad, las condiciones materiales de vida y la integridad física y moral. En el Derecho español, la jurisprudencia constitucional concibe la dignidad desde una perspectiva axiológica como un «valor espiritual y moral inherente a la persona, que se manifiesta singularmente en la autodeterminación consciente y responsable de la propia vida y que lleva consigo la pretensión al respeto por parte de los demás» (STC 53/1985 de 11 de abril), situándola en el fundamento, entre otros, de los derechos a la integridad física y moral, la libertad ideológica, el honor y la intimidad personal, la propia imagen, la libertad informativa y la protección de los datos personales.

En todos esos ámbitos jurídicos pueden producirse daños como consecuencia de la interacción con la robótica y la inteligencia artificial, y la dignidad humana respaldará, como fundamento último, las medidas que el ordenamiento jurídico debe imponer para evitar aquellos daños y atender a su reparación.

Medida jurídica esencial de esta puesta en práctica del principio de no maleficencia será el consentimiento informado que debe preceder a la interacción con sistemas robóticos o de inteligencia artificial que pueden producir daños. Desde luego, en el ámbito médico y de la investigación es ya una exigencia legal, que será inmediata y directamente aplicable a los sistemas robóticos y de inteligencia artificial que se desenvuelvan en esos ámbitos, como también ocurre en el tratamiento de datos de carácter personal, y estará llamada a ser una regla general en el caso de que la personalidad de aquellos sistemas sea una realidad jurídica.

Por último, la prevención de daños a la que responde el principio de no maleficencia cobra un valor especial al destinarlo a grupos demográficos vulnerables. Así, el Borrador de directrices éticas destaca que

[...] los niños, las minorías, las personas discapacitadas, las personas de edad avanzada o los inmigrantes deberían recibir mayor atención a la prevención de los daños, dada su condición única en la sociedad. La inclusión y la diversidad son ingredientes clave para la prevención de los daños, a fin de garantizar la idoneidad de estos sistemas en todas las culturas, géneros, edades, opciones de vida, etc. Por lo tanto, la inteligencia artificial no solo debe diseñarse teniendo en cuenta el impacto en diversos grupos demográficos vulnerables, sino que los grupos demográficos mencionados anteriormente deben tener un lugar en el proceso de diseño.

4.3.1.2. *Beneficencia. Procurar el bien*

Al vincular este principio bioético con la inteligencia artificial, el Borrador de directrices éticas destaca que estos sistemas deben ser diseñados y desarrollados para mejorar el bienestar individual y colectivo, y los eventuales beneficios derivados de la robótica y de la inteligencia artificial se traducen en

[...] prosperidad, creación de valor y maximización y sostenibilidad de la riqueza [...] la protección de los procesos democráticos y el estado de derecho; la provisión de bienes y servicios comunes a bajo costo y de alta calidad; la alfabetización y la representatividad; [...] el logro de los Objetivos de Desarrollo Sostenible de la ONU o la sostenibilidad entendida de manera más amplia, de acuerdo con los pilares del desarrollo económico, la equidad social y la protección del medio ambiente.

El principio de beneficencia responde a la afirmación general de hacer el bien en el sentido más amplio, y se concibe como la obligación moral de actuar en beneficio de otros. Tiene un alcance más reducido que el principio de no maleficencia, pues aunque existe una obligación general y constante de no perjudicar a los demás, no siempre existirá el deber de beneficiarlos. Por lo general se ha considerado el principio más importante en el ámbito de la bioética, pero hoy se entiende modalizado por el principio de autonomía, abandonando así el paternalismo cercenador del libre desarrollo de la personalidad del individuo a costa de proporcionarle un beneficio no consentido.

Claro que en la valoración del alcance de los beneficios derivados de la robótica no juegan solo los eventualmente obtenidos por el usuario concreto del robot o del sistema de que se trate, sino que la cuestión enlaza, como ocurre en general con las innovaciones tecnológicas encaminadas a la mejora indi-

vidual, con el principio de justicia y el debate acerca de la necesaria distribución igualitaria de dichos beneficios.

Y así se ha sostenido que una distribución desigualitaria de los beneficios derivados de la robótica podría (Asis, 2013)

(i) provocar una sociedad compuesta por individuos de primera (que pueden acceder a la nueva tecnología) y de segunda (que no pueden acceder a la nueva tecnología y que lógicamente serán dominados por los primeros), en la que además [...] se eliminaría [...] la competitividad social y la diversidad (ya que el esfuerzo y la diversidad son sustituidos por máquinas). (ii) [...] alterar la identidad de las personas y transgredir su privacidad; (iii) [...] provocar desde un punto de vista global el exterminio de otras razas y culturas y el sometimiento de otros pueblos; (iv) [...] y [...] apropiarse de un supuesto derecho a tomar decisiones que comprometan el futuro.

Claro que, frente a ello, puede afirmarse que la desigualdad está más cerca de ser una constante histórica que una novedad provocada por las nuevas tecnologías y que, al contrario que lo sostenido por aquellas posiciones doctrinales, la robótica y la inteligencia artificial pueden ser un mecanismo idóneo para reducir la desigualdad, por ejemplo, en el ámbito de las discapacidades físicas y psíquicas. Y justo en ese ámbito cobra también relevancia, a la hora de analizar los eventuales beneficios derivados de la robótica y la inteligencia artificial desde el punto de vista ético, la distinción entre intervenciones terapéuticas e intervenciones que implican mejora.

Un ejemplo puede servir para introducir esa distinción: un hipotético sistema de inteligencia artificial permite a los discapacitados cognitivos compensar sus deficiencias y atribuirles una capacidad intelectual equiparable al nivel social medio; dotado de personalidad jurídica, dicho sistema podría configurarse jurídicamente como un cargo tutelar algorítmico y facultaría a los discapacitados para «gobernarse por sí mismos» (Código civil español, artículo 199) y desenvolverse con autonomía, impidiendo que sus propios actos puedan ocasionar perjuicios tanto a ellos como a terceras personas.

Tributaria del principio de igualdad, la posibilidad de permitir a los discapacitados cognitivos acceder, respaldados por un sistema de inteligencia artificial, a una vida autónoma por haber alcanzado el nivel intelectual que la sociedad considera estándar, no debería plantear mayor problema ético. Sin embargo, la atribución de capacidades intelectuales superiores a lo considerado normal excede, para muchos, de lo que el principio de beneficencia puede amparar.

No es objeto de este trabajo el estudio de la distinción entre intervenciones curativas o paliativas, por un lado, y las de mejora, por otro, aunque merecen destacarse dos criterios contrapuestos para delimitar el contenido estándar a partir del cual toda intervención cuantitativa o cualitativa se considera mejora y que, al contrario, cuando no llega a colmarse, reclama medidas terapéuticas.

Así, un criterio comparativo o relativista de las competencias humanas sitúa la nota distintiva en la «diferencia relativa» entre las habilidades o facultades de las personas promedio y las de las personas que, por exceso o por defecto, se alejan de la media; dicha diferencia se traducirá en diversos resultados en el ejercicio de esas habilidades, y ese resultado repercutirá no solo en el individuo, sino también en la sociedad. La capacidad humana innata para categorizar permite percibir a los demás y a uno mismo como integrantes de grupos en atención al nivel de competencia en diversos campos psíquicos y físicos, y esa categorización es trascendental para la organización social: la incompetencia psíquica que impide al sujeto gobernarse por sí mismo conduce a una privación del ejercicio de sus derechos, función que se encomienda a un cargo tutelar, de igual modo que desde el punto de vista físico, por naturaleza o por accidente, cualquier ser humano puede estar por debajo del promedio. En el otro extremo, los genios intelectuales y los atletas exceden con mucho del promedio competencial humano. Esta idea se ha propuesto (Wikler: 2017, 360) en relación con la capacidad intelectual, pero bien puede utilizarse, en general, con respecto al principio de beneficencia y al estudio del mejoramiento humano.

Siguiendo esta idea, desde la óptica de John Rawls (*cit.* Wikler: 2017, 360), se sostiene que la competencia natural humana es un bien común del que disponen de forma igualitaria todos los seres humanos, de modo que la competencia se determina, ya no por comparación entre individuos, sino a partir de un estándar absoluto que habilita para el ejercicio normal de la facultad de que se trate.

Estos dos sistemas de valoración de la competencia humana, atendiendo a un criterio relativista (comparativo), o considerándola un bien común, se han propuesto (Wikler: 2017) en relación con la discapacidad intelectual, pero bien pueden utilizarse con un carácter general para comprender las diversas opiniones relacionadas con la ética del mejoramiento y su relación con el principio de beneficencia y el paternalismo.

Siguiendo un esquema relativista, resulta más fácil justificar como terapéuticas o paliativas las intervenciones que, superando el promedio, atribuyen al sujeto competencias o habilidades que le permiten sobresalir de la media, lo que explica el éxito comercial de fármacos nootrópicos o del sinedafil. De igual modo, en el ámbito de la capacidad intelectual, el relativismo comparativo

impone aceptar el sometimiento al paternalismo, cuando la diferencia de competencia entre los sujetos normales y los mejorados sea lo bastante perceptible (y cuantitativamente equiparable a la que separaba antes a los primeros de aquellos que no alcanzaban el promedio), y admitir así, que los sujetos, antes normales (y que hayan optado por no mejorarse), puedan quedar privados del ejercicio de sus derechos y sometidos a un sistema tutelar.

La óptica de la competencia como un bien común implica el establecimiento de un umbral en función de la facultad cuyo ejercicio se trata de valorar (Wikler: 2017, 361), por lo que, a la hora de ponderar la discapacidad, parece tutelar mejor los derechos de los sujetos no mejorados en caso de que se generalizasen, como mejoramiento, tecnologías de inteligencia artificial y «no justificaría tan fácilmente el paternalismo como el enfoque relativista puede parecer». Ya no se trata de que el grupo de mejorados sea de manera objetiva más competente que los sujetos normales, sino de que estos sigan disponiendo de actitudes que les permitan «gobernarse por sí mismos» y decidir si desean o no ser mejorados; aquel hipotético sistema de inteligencia artificial no actuaría en los sujetos normales como un cargo tutelar, sino que más bien le correspondería el papel de un asesor o consejero, pues, al fin y al cabo, al carecer de las ventajas proporcionadas por la mejora, serían menos competentes que los sujetos mejorados.

4.3.1.3. *Autonomía. Consentimiento informado*

En su acepción bioética primigenia, en el ámbito de la bioética médica, el principio de autonomía surgió como contrapeso a la concepción paternalista tradicional que justificaba la imposición de un beneficio al sujeto sin contar con su voluntad, e incluso en contra de esta.

En su acepción más amplia, es una manifestación de la libertad humana, y en la Resolución del Parlamento Europeo se concibe como la capacidad de tomar una decisión independiente y con conocimiento de causa sobre los términos de interacción con los robots. En cierto modo, ya aparecía reflejado en la tercera de las leyes de Asimov («Un robot debe cumplir las órdenes dadas por los seres humanos, a excepción de aquellas que entren en conflicto con la primera ley»), que lo contemplaba en su versión negativa; a nivel jurídico, su vertiente positiva descansa directamente en la dignidad y la autodeterminación humanas, e implica la necesidad de poner en conocimiento de los ciudadanos el hecho de que van a interactuar con un sistema de inteligencia artificial o con un robot a fin de que puedan decidir con conocimiento de causa.

Esa fundamentación jurídica del principio de autonomía se traduce, desde el punto de vista técnico, en el necesario control humano (*human-in-command*) sobre los robots y la inteligencia artificial, de modo que los humanos conserven en todo momento el dominio sobre ellos, tal y como se expuso (*supra* 4.4.7.4.) al tratar del Dictamen del Comité Económico y social.

Todo ello es reflejo del antropocentrismo, acogido de forma expresa en el Borrador de directrices éticas como uno de sus pilares, lo que obliga a considerar la inteligencia artificial y, por extensión, la robótica, no como un fin en sí mismo, sino como un medio para asegurar que «los valores humanos son la consideración primordial [...] con el objetivo de aumentar el bienestar de los ciudadanos».

Requisito esencial del principio de autonomía es la capacidad de obrar del sujeto, y premisa para su ejercicio será la información adecuada que permita al sujeto tomar conocimiento de los eventuales riesgos y beneficios de su interacción con un sistema robótico o de inteligencia artificial. Todo ello requerirá un sistema normativo que reduzca en lo posible las asimetrías de información y poder, especialmente en las relaciones empresa-consumidor y administración pública-ciudadano.

Por su parte, Natalie Nevejans destaca que de este principio de autonomía derivan las siguientes ideas:

- El ser humano debería tener siempre la capacidad de obligar a un robot a obedecer sus órdenes y debería excluirse su utilización como armas.

A esta última cuestión se refiere de forma expresa la Resolución del Parlamento europeo 2108/2752 del 5 de septiembre, que insta a la ONU y a la OTAN a que promuevan la prohibición «del desarrollo, la producción y la utilización de sistemas armamentísticos totalmente autónomos que permitan realizar ataques sin intervención humana significativa» y reclama la necesidad de «un instrumento jurídicamente vinculante que prohíba los sistemas armamentísticos completamente autónomos». Como contrapunto a esta medida elemental de protección de la dignidad y la libertad humanas, Estados Unidos, Israel, Rusia, Reino Unido y Corea del Sur están desarrollando sistemas armamentísticos autónomos que, sin duda, facilitarán el inicio de los conflictos bélicos y su desplazamiento hacia la población civil abriendo las puertas, según el Borrador de directrices éticas, a una carrera armamentística incontrolable sin precedentes históricos. En ambos textos se reclama que la responsabilidad debe recaer siempre sobre los seres humanos.

- Es necesario adoptar precauciones especiales en la interacción de sistemas de inteligencia artificial y robots con personas especialmente vulnerables (niños, ancianos, discapacitados).

Ya se ha expuesto que el principio de autonomía presupone en su ejercicio el requisito de la capacidad de obrar, de modo que respecto a menores e incapacitados entrarán en juego todos los mecanismos jurídicos tuitivos dispuestos por el ordenamiento jurídico, reconduciendo, en su caso, a sus representantes legales la prestación del consentimiento informado para la interacción con sistemas robóticos o de inteligencia artificial. Respecto a los ancianos, será decisivo equilibrar los innegables beneficios que para su autonomía están llamados a proporcionar los sistemas robóticos asistenciales y los eventuales riesgos, especialmente los derivados de la vulneración de su intimidad y de la protección de sus datos personales.

Al mismo tiempo, respecto a las personas en especial vulnerables e incapacitadas de manera legal, es necesario plantearse, a la hora de prestar su consentimiento (menores de edad o incapacitados judicialmente), hasta qué punto ciertas aplicaciones robóticas terapéuticas podrían tener un carácter coercitivo en contra de la opinión de sus representantes legales. Es evidente que el principio jurídico que debe inspirar la solución que se adoptaría sea el del mayor beneficio para la calidad de vida del sujeto (tal y como ya proclama, en general, en el ámbito de la salud el artículo 9.6 de la Ley 41/2002 de Autonomía del paciente), abriendo así la posibilidad del recurso a la tutela judicial ante negativas de los padres o representantes legales a la implantación de soluciones robóticas o de inteligencia artificial que, objetivamente, supongan una mejora terapéutica. La negativa a soluciones terapéuticas tecnológicas no es desde luego una novedad; se ha planteado ya en el caso de los implantes auditivos, rechazados por algunos padres alegando que su uso excluiría a sus hijos de la cultura sorda. Las perspectivas terapéuticas que se abren de la mano de la robótica exigen a la sociedad que esté preparada para evitar que prejuicios culturales, religiosos o ideológicos priven a las personas vulnerables del acceso a mejoras en su salud y calidad de vida.

La separación entre la legítima integración de los menores en la cultura propia de los padres y la privación del acceso a un recurso terapéutico es paralela a la distinción entre cualidades subjetivas y capacidades objetivas. Las primeras pueden, en su caso, hacer sufrir a los sujetos portadores de esta discriminación social (color de piel, religión), mientras que la falta de las segundas origina una carencia sensorial objetiva. Confundir integración cultural y discapacidad (Hervás 2001: 136) conduce

[...] al absurdo de considerar la curación, y la recuperación de las capacidades inexistentes, como la eliminación de lo que identifica a la persona, casi equivalente a su muerte, cuando por el contrario el sentido común nos invita a pensar, como señala ROCHE, que la restauración de la vista o la motricidad sería una mejora y no una violación de la identidad de la persona. Incluso si se acepta la consideración de la discapacidad como un signo de identidad que permite incluir a esta persona como miembro de una comunidad de pares, con una cultura y valores propios, aún habría que justificar por qué ha de valorarse la pertenencia a la misma, y lo que es más, que dicha comunidad tenga fundamentos para requerir a individuos que no pertenecen a la misma que contribuyan a aportar nuevos miembros.

- Debe exigirse el consentimiento previo y expreso del paciente para la prestación de servicios sanitarios por parte de robots.

Se trataría de una manifestación específica del principio general del consentimiento médico informado y del derecho de negarse a recibir o desistir de determinado tratamiento médico. Complemento de ese derecho a rechazar un servicio sanitario prestado por un sistema robótico es el derecho del paciente a la elección de terapia, concretado en nuestro derecho positivo en el artículo 3 de la Ley 41/2002 básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica como libre elección, es decir, la facultad del paciente o usuario de optar, libre y voluntariamente, entre dos o más alternativas asistenciales. Ocurre que dicha libre elección en el ámbito de la sanidad pública quedará limitada por razones económicas y de eficiencia, de modo que si en la Seguridad Social llegase a implantarse como universal un determinado procedimiento robótico en la práctica médica (por ejemplo, un sistema robótico autónomo de extracción y análisis de sangre), sería muy difícil defender la exigencia de un procedimiento alternativo con intervención humana.

- Todo ser humano debe tener derecho a negarse a recibir atención de un robot, incluso aunque no exista la posibilidad de sufrir ningún daño.

Esta afirmación, referida de nuevo al ámbito de los cuidados médicos y asistenciales, se fundamenta, según el Estudio del Parlamento Europeo, en que aun siendo inexistente el daño, la dignidad humana podría justificar la negativa si el paciente «se siente tan profundamente incómodo en presencia del robot que tal presencia le resulta insoportable». Como ejemplo se cita el caso de los robots encargados de alimentar a las personas mayores o

discapacitadas. De nuevo esta cuestión se relaciona con la libertad de elección de tratamiento ya mencionada.

Por su parte, el Borrador de directrices éticas afirma que el consentimiento informado «es un valor necesario para hacer operativo en la práctica el principio de autonomía», pues, argumenta, los seres humanos deben mantener una autodeterminación plena y efectiva sobre sí mismos, y esto implica el reconocimiento de los derechos a ser previamente informados de la interacción con un sistema de inteligencia artificial, de exclusión voluntaria y de desistimiento. Lo anterior no será obstáculo para que, en determinados supuestos, dicho consentimiento no sea preciso para la interacción con sistemas robóticos o de inteligencia artificial, por ejemplo, de modo similar a lo que ocurre en materia de protección de datos de carácter personal cuando esa interacción sea necesaria para proteger un interés vital del interesado o de un tercero, cuando exista habilitación legal para ello, o cuando sea precisa para el cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos (Reglamento UE 2016/679 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a su libre circulación).

Ahora bien, aun cuando el consentimiento para la interacción con sistemas robóticos pueda llegar a ser prescindible en algunos casos, siempre debería ser necesario el conocimiento por parte del usuario, excluyendo la posibilidad de sistemas robóticos o de inteligencia artificial encubiertos. En esta línea, el Borrador de directrices éticas señala que

[...] un humano siempre tiene que saber si está interactuando con un ser humano o con una máquina, y es responsabilidad de los desarrolladores [...] que esto se logre de forma fiable. De lo contrario, las personas con el poder de controlar (los sistemas robóticos o de inteligencia artificial) serán potencialmente capaces de manipular a los humanos a una escala sin precedentes.

En especial problemáticos serán los casos fronterizos como el de voces humanas filtradas por sistemas de inteligencia artificial, realidad aumentada generada por inteligencia artificial o sistemas híbridos biológicos-robóticos.

4.3.1.4. Justicia. Acceso igualitario a los beneficios de la robótica

La distribución de recursos (por naturaleza limitados) y la igualdad (consagrada como principio constitucional) son los dos aspectos, respectivamente, material y formal, del principio bioético de justicia.

En su aplicación concreta a la robótica y a la inteligencia artificial, el principio de justicia puede relacionarse con las siguientes cuestiones:

- la distribución justa de los beneficios asociados a la robótica;
- la asequibilidad de los robots y de los sistemas de inteligencia artificial;
- la exclusión del «sesgo algorítmico», evitando prejuicios, estigmatización y discriminación en la implementación de estos sistemas;
- la implantación de mecanismos de reparación de los daños sociales e individuales ocasionados por estas tecnologías.

Es un lugar común destacar las notables dificultades que ofrece la puesta en práctica del principio de justicia distributiva en los sectores dominados por la iniciativa empresarial privada, obstáculo al que, en los campos de la robótica y de la inteligencia artificial, se añade el hecho de que sus posibles aplicaciones y finalidades son tan diversas que rechazan cualquier intento de generalización. En efecto, no se puede pretender la misma relevancia del principio de justicia distributiva respecto a la implantación de robots destinados a la asistencia geriátrica o de discapacitados que de aquellos destinados a prácticas sexuales, pero la variedad de usos es tal que ni siquiera en este último sector puede generalizarse; piénsese, por ejemplo, en la asistencia sexual robótica a personas con diversidad funcional.

Desde luego, al tratarse de sistemas robóticos o de inteligencia artificial empleados en los ámbitos médico y asistencial, son perfectamente trasladables a estos campos los planteamientos y criterios consolidados en el sector de la bioética médica, y es que, como destaca el Proyecto de informe, una «brecha robótica» sería en especial preocupante en los sectores sanitario y geriátrico, lo que reclama vías de financiación públicas (ventajas fiscales, de cotización, subvenciones, etc.) a fin de garantizar el acceso en igualdad de condiciones a los sectores sociales más desfavorecidos. Y otro tanto puede afirmarse del sector educativo y de protección de los discapacitados.

En estos aspectos económicos del principio de justicia, tendrá especial relevancia la regulación de los derechos de propiedad industrial e intelectual de los sistemas robóticos y de inteligencia artificial, en los que pueden servir de pauta, en los referidos ámbitos (sanitario y asistencial), los principios inspiradores de la Declaración de Doha, que reconoció que los medicamentos y las

tecnologías sanitarias no son como las demás mercancías, y deben diferenciarse de otras invenciones, de modo que los acuerdos internacionales en materia de patentes no deben impedir que los Estados puedan adoptar medidas para proteger la salud pública y promover el acceso a los fármacos, tan necesario para toda la población. Licencias obligatorias, plazos reducidos para ciertas patentes, exigencia de código abierto, entre otras cosas, son medidas coherentes con el principio de justicia distributiva que bien podrían imponerse como contrapartida al beneficio que para fabricantes y titulares supondría la eventual concesión de personalidad jurídica a los robots.

Además de estos aspectos económicos, el principio de justicia también está llamado a tutelar otros de carácter eminentemente jurídico y político, evitando la discriminación y los prejuicios hacia grupos sociales minoritarios y/o vulnerables. La pretensión ilusoria de que la naturaleza abstracta y objetiva de las matemáticas garantizaría la igualdad y la justicia de las decisiones basadas en el tratamiento algorítmico del *big data* se da de bruces con la realidad cuando se observa que los sistemas automatizados de decisión reproducen los mismos prejuicios, arbitrariedades y sesgos de que adolecen las decisiones tomadas por los seres humanos.

Contratación laboral, calificación crediticia y académica, prevención delictual, entre otros, son sectores bastante sensibles, aunque podría argumentarse que los prejuicios y la estigmatización siempre han existido y que, por lo tanto, las decisiones de los sistemas de inteligencia artificial no podrían ser más injustas o discriminatorias que las tomadas por seres humanos. Sin embargo (O'Neil, 2017: 250-252), existe una notable diferencia, pues los procesos de *big data* codifican hechos pasados, de modo que permanecen congelados en el tiempo, impermeables a eventuales cambios éticos de la sociedad. Además, al ser diseñados para maximizar beneficios, son el único criterio que tomarán en consideración las empresas que los utilizan y así (O'Neil, 2017: 251)

[...] cuando un programa informático identifica con éxito a las personas que están suficientemente desesperadas como para pagar un tipo de interés del 18 % al mes, los que recogen los beneficios piensan que el programa funciona a la perfección [...]. Como es obvio, los prestamistas de crédito rápidos y otras empresas de su índole irán primero a por los pobres y los inmigrantes. Son los blancos más fáciles, los objetivos más cercanos. Son personas que tienen menos acceso a la información.

Para evitar estos inconvenientes, el Borrador de directrices éticas recuerda a los programadores e implementadores de sistemas de inteligencia artificial la

vigencia del principio de justicia y su deber de asegurar que los «individuos y los grupos minoritarios se mantengan libres de prejuicios, estigmatización y discriminación [...] evitando colocar a la población vulnerable en una situación de mayor vulnerabilidad». Y es de esperar que, en respuesta a esta llamada y antes de la aparición de normas jurídicas reguladoras de estas cuestiones, surjan mecanismos de autorregulación similares a *The financial modeler's manifesto*, propuesto, a modo de juramento hipocrático de los programadores de algoritmos, por Enmanuel Derman y Paul Willmott como una respuesta a la crisis que siguió al colapso del sistema de las hipotecas de alto riesgo en Estados Unidos.

4.3.1.5. Precaución. Controlar el riesgo

La Resolución del Parlamento Europeo somete la investigación y el desarrollo de los sistemas robóticos y de inteligencia artificial al principio de precaución, lo que exige anticiparse «a los posibles impactos sobre la seguridad y adoptar las medidas debidas, en función del nivel de protección, al tiempo que se fomenta el progreso en beneficio de la sociedad y del medio ambiente».

Esta llamada al principio de precaución es deudora del artículo 194 del Tratado de Funcionamiento de la Unión Europea, que somete a «los principios de cautela y de acción preventiva» cualquier política de la Unión Europea en materia de medio ambiente. Formulado en origen en este ámbito jurídico determinado, el principio de precaución se concibe actualmente como un principio general del Derecho de la Unión que puede aplicarse a todas aquellas actividades que pueden poner en riesgo la seguridad y la salud humana, animal y vegetal.

Según la Comunicación de la Comisión suscrita en Niza en el año 2000, el principio de cautela solo debe actuar en el ámbito normativo en caso de «riesgo potencial, aunque este riesgo no pueda demostrarse por completo, no pueda cuantificarse su amplitud o no puedan determinarse sus efectos debido a la insuficiencia o al carácter no concluyente de los datos científicos» y, dado que su aplicación se traducirá en normas prohibitivas o restrictivas, no debe tolerarse su puesta en práctica arbitraria e indiscriminada.

El primer requisito al que se somete la aplicación de este principio es la adecuada evaluación de los datos científicos sobre los riesgos, lo que en la práctica legislativa de la Unión Europea se traduce en cuatro fases: identificación del peligro, caracterización del peligro, evaluación de la exposición y caracterización del riesgo. Por supuesto, el recurso al método científico no excluye, más

bien al contrario, impone, la entrada en juego de la incertidumbre y de márgenes de error, y, así, la puesta en práctica del principio de precaución presupone, en el Derecho de la Unión Europea:

- La identificación de efectos potencialmente peligrosos que se derivan de un fenómeno, de un producto o de un proceso y
- una evaluación científica de los riesgos que, debido a la insuficiencia de los datos, a su carácter no concluyente o a su imprecisión, no permite determinar con una certeza suficiente el riesgo en cuestión.

Ante tal situación, el poder político debe determinar cuál es el nivel de riesgo aceptable por la sociedad (en el ámbito de los contaminantes tóxicos se alude al nivel «ALARA», *as low as reasonably achievable*, «lo más bajo que sea razonablemente posible»), y a partir de ahí, tomar la decisión pertinente, teniendo en cuenta que «el principio de precaución no se traduce necesariamente en la adopción de actos finales destinados a producir efectos jurídicos que pueden ser objeto de un control jurisdiccional» y que, por el contrario, las posibilidades de actuación política que ofrece tal principio son muy variadas y pueden ser de orden económico (financiar, o no, determinadas líneas de investigación, conceder ciertas subvenciones o beneficios fiscales, etc.), social (informar a la opinión pública sobre ciertos riesgos o beneficios, entre otras cosas) o legislativo (prohibir o regular determinadas actividades, etc.); incluso la inacción es una decisión posible.

Según el criterio de la Unión Europea, la aplicación del principio de precaución obliga a la ponderación de los costes

[...] de la acción y de la inacción, tanto a corto como a largo plazo, lo que no se limita sencillamente a un análisis económico de rentabilidad, sino que abarca un ámbito mucho más amplio e incluye consideraciones no económicas, como la eficacia de las posibles opciones y su aceptabilidad para la población. Al llevar a cabo tal examen, deberá tenerse en cuenta el principio general y la jurisprudencia del Tribunal de que la protección de la salud tiene prioridad sobre las consideraciones económicas.

Trasladar estas consideraciones al escenario planteado por el eventual reconocimiento de personalidad jurídica a los robots y sistemas de inteligencia obliga a determinar los posibles perjuicios que se derivarían de dicha personificación, más allá de las alegaciones genéricas a la quiebra de los pilares humanistas de nuestra civilización y deslindar claramente los riesgos específicos que

podría originar aquel reconocimiento subjetivo de aquellos otros inherentes, en general, a la expansión de la robótica y de la inteligencia artificial.

4.3.1.6. Transparencia. Contra la opacidad del algoritmo

El principio de transparencia aparece consagrado en las disposiciones de aplicación general en el artículo 15 del Tratado de Funcionamiento de la Unión Europea, imponiendo a todas las instituciones, órganos y organismos de la Unión actuar «con el mayor respeto posible al principio de apertura» garantizando la transparencia en todas sus actuaciones.

También en el ámbito de nuestra normativa interna se acoge la transparencia como uno de los principios generales del Derecho administrativo moderno, y así lo proclaman la Ley 39/2015 del 1 de octubre de Procedimiento Administrativo Común de las Administraciones Públicas y la Ley 19/2013 del 9 de diciembre de transparencia, acceso a la información pública y buen gobierno.

Además, y en el ámbito de las relaciones electrónicas entre las administraciones públicas, el artículo 157 de la Ley 40/2015 del 1 de octubre del Régimen jurídico del sector público regula la puesta a disposición de aquellas de todas «las aplicaciones, desarrolladas por sus servicios o que hayan sido objeto de contratación y de cuyos derechos de propiedad intelectual sean titulares» y dispone que tales aplicaciones puedan ser «declaradas como de fuentes abiertas, cuando de ello se derive una mayor transparencia en el funcionamiento de la Administración Pública o se fomente con ello la incorporación de los ciudadanos a la Sociedad de la información». Además, el artículo 41 de la misma Ley, al regular las actuaciones administrativas automatizadas (aquellas realizadas íntegramente a través de medios electrónicos por una administración pública y en las que no haya intervenido de forma directa un empleado público), dispone que, previamente, deberá establecerse el órgano u órganos competentes para la definición de las especificaciones, programación, auditoría y control del sistema y de su código fuente.

El principio de transparencia está también llamado a desempeñar un papel esencial en la futura regulación europea sobre inteligencia artificial, y así, según el Proyecto y la Resolución del Parlamento Europeo sobre robótica,

[...] siempre ha de ser posible justificar cualquier decisión que se haya adoptado con ayuda de la inteligencia artificial y que pueda tener un impacto significativo sobre la vida de una o varias personas [...] [y] siempre debe ser posible

reducir los cálculos del sistema de inteligencia artificial a una forma comprensible para los humanos.

Para ello, propone un entorno abierto, que abarque normas abiertas, modelos de licencia innovadores y plataformas abiertas, con el fin de evitar el confinamiento en sistemas privados que limiten la interoperabilidad.

Por su parte, el Borrador de directrices éticas para una inteligencia artificial confiable del grupo de expertos de la Unión Europea propone la aplicación del principio de transparencia no solo al modelo tecnológico en sí, sino también al modelo de negocio. Referida al aspecto tecnológico, la transparencia implica que los sistemas de inteligencia artificial sean «auditables, comprensibles e inteligibles»; con respecto al modelo de negocio, la transparencia exige que los ciudadanos estén informados «de la intención de los desarrolladores e implementadores» de estas tecnologías. En ambos aspectos, el principio de transparencia es, en cierto modo, tributario del principio de no maleficencia y del consentimiento informado y, además, impone que quede constancia de todas las decisiones que tomen los sistemas de inteligencia artificial, así como del proceso que dio lugar a cada una de esas decisiones, garantizando, en suma, su trazabilidad, pues

[...] si bien la trazabilidad no siempre puede decirnos por qué se tomó una determinada decisión, sí puede decirnos cómo se tomó, lo que permite razonar por qué una decisión de IA fue errónea y puede ayudar a prevenir futuros errores. La trazabilidad es, por lo tanto, un facilitador de la auditabilidad, lo que implica la habilitación y facilitación de la monitorización y verificación de algoritmos, datos y procesos de diseño.

4.3.2. Leyes de Asimov para humanos o Derecho natural para robots

Es interesante destacar que, al plantear la regulación de los aspectos éticos de la robótica y la inteligencia artificial, tanto el Proyecto de informe como la Resolución del Parlamento Europeo, reiteran que

[...] debe entenderse que las leyes de Asimov van dirigidas a los diseñadores, fabricantes y operadores de robots, incluidos los que disponen de autonomía y capacidad de autoaprendizaje integradas, dado que dichas leyes no pueden traducirse en código máquina.

Concretando dicha proclamación de principios, la Resolución, además de las propuestas de orden normativo que realiza, contiene, en su anexo, códigos de conducta dirigidos a los ingenieros en robótica y a los comités de ética de la investigación para la revisión de los protocolos de robótica. Se reconoce, pues, de forma explícita, la dificultad de programar aquellos principios a modo de Derecho natural para robots, pero ello no puede equivaler a una renuncia a implantar reglas morales en el código máquina de los sistemas de inteligencia artificial.

Ciertamente, aquella dificultad obedece tanto (Monasterio, 2018: 136-140) a razones teóricas (no hay ninguna teoría ética que recoja de manera suficiente y comprehensiva la complejidad del fenómeno moral ni hay acuerdo a la hora de definir los valores —ni los principios éticos que los representan— e identificar los mecanismos o procesos genéticos, neuronales, psicológicos y culturales implicados en la toma de decisiones morales) como a razones prácticas y metodológicas (referentes a la implementación concreta del razonamiento ético en sistemas artificiales y al diálogo necesario entre programadores y expertos en ética), pero no debe entenderse como imposibilidad y renuncia a la implementación de una ética para máquinas. En efecto, la distinción entre juegos de información completa (aquellos en los que, como el go o el ajedrez, toda la información necesaria está a disposición de los jugadores) y juegos de información incompleta (como el hanabi, donde es necesario tener en cuenta la creencias e intenciones de otros jugadores) permite una aproximación al desenvolvimiento de sistemas de inteligencia artificial en entornos complejos como las relaciones sociales con juicios morales. Así se han entrenado redes neuronales artificiales para ejercitar una teoría de la mente (conocidas como ToMnet) y son capaces de pasar con éxito el denominado test de Sally-Anne, que obliga a tener en cuenta los estados mentales de otros sujetos y, así, y mediante el

[...] uso de algoritmos de aprendizaje por refuerzo autónomo y juegos concebidos como entornos de entrenamiento de los algoritmos para que aprendan con la práctica, compañías como DeepMind están aproximándose a la tarea de conseguir inteligencia a nivel humano con todo lo que conlleva, incluida la competencia moral, de una manera orgánica, desde cero, o dicho de otra manera: desde un enfoque abajo-arriba. Pero también dentro de la ética para máquinas cuyo objetivo último es conseguir producir máquinas morales existe otro enfoque que busca programar reglas o meta-reglas de manera inicial para producir comportamiento moral. Este es el conocido como enfoque arriba-abajo.

Admitida la impronta como un mecanismo biológico de aprendizaje con base genética (sin duda, es una ventaja evolutiva que las crías de las aves y rep-

tiles reconozcan como progenitor al primer ser mayor que ellos que ven al romper el cascarón, pues lo será en la gran mayoría de los casos, y solo en un porcentaje de supuestos estadísticamente irrelevante, al romper el cascarón, las crías se encontrarán con un depredador o con un científico realizando estudios sobre la impronta y el apego filial), se ha visto también alguna aplicación biológica de esta con trascendencia moral en seres humanos (tal es la propuesta de Edvard Alexander Westermarck para explicar el incesto como un tabú universal). Trasladar esas ideas al ámbito de los sistemas de inteligencia artificial, tanto implícitos (recuérdese, el cajero automático que debe impedir el acceso a cuentas distintas de la del cliente que está operando) como explícitos (el coche automático que debe decidir la opción del menor daño posible en caso de accidente), o incluso plenos, nos situaría ante una verdadera ética integrada en el código de la máquina. ¿Un Derecho natural robótico?

4.3.3. Contenido de la personalidad algorítmica y requisitos constitutivos

Ni el Proyecto de informe ni la Resolución del Parlamento Europeo concretan el contenido de la eventual personalidad de los robots y sistemas de inteligencia artificial. El Proyecto se limita a afirmar que dicha personalidad atribuiría «derechos y obligaciones específicos, incluida la obligación de reparar los daños que puedan causar», y la Resolución, que se trataría de una «personalidad específica». Desde luego, queda descartada la posibilidad de atribución de derechos fundamentales inherentes a la dignidad humana, no solo por incongruente, sino por innecesario, tal y como ocurre con los derechos a la vida, libre desarrollo de la personalidad, integridad física, estado civil, libertad ideológica, salud, etc., que tampoco corresponden a las personas jurídicas.

Lo anterior no es obstáculo para que tales personas no humanas, en la actualidad las jurídicas y, eventualmente, las algorítmicas o electrónicas, puedan ostentar otros derechos de rango constitucional. Al fin y al cabo, tanto las unas como, en algunas ocasiones, las otras nacen a la vida jurídica en ejercicio de los derechos fundamentales del ser humano: el de asociación (consagrado en el artículo 22 de la Constitución española), en el caso de la persona jurídica, y el de libre investigación y producción científica y técnica (ex artículo 20), en el caso de esa eventual personalidad algorítmica o electrónica. Y es que la determinación concreta de los derechos que correspondan a cada categoría de sujeto de derecho es algo que resuelve cada ordenamiento jurídico, pues, al ser la personalidad forma dada por el Derecho,

[...] cabe la posibilidad de que esta personalidad sea atribuida en más o menos larga medida para ciertos respetos y no para otros [...]. Por motivos de política legislativa, el orden jurídico da un contenido más bien que otro a la configuración unitaria de la personalidad (Ferrara, 2006: 258).

De esta forma puede hacerse un recorrido por los derechos, algunos de rango constitucional, que podrían corresponder a los sistemas de inteligencia artificial con personalidad jurídica:

- La nacionalidad

Correspondería a los sistemas de inteligencia artificial con personalidad jurídica, no en su vertiente de derecho fundamental, sino en su concepción como vínculo jurídico con el estado y sometimiento a un ordenamiento jurídico regulador una vez cumplidos los requisitos exigidos por ese ordenamiento para el surgimiento de la personalidad jurídica. A la hora de determinar el punto de conexión que atribuya la nacionalidad frente a criterios como la sede física, el lugar donde el sistema desarrolle su actividad, o el lugar de matrícula o registro, desempeñará un papel relevante en el lugar donde se ubiquen los servidores que alojen el *software* del robot o sistema de inteligencia artificial.

- El domicilio y el secreto de las comunicaciones

Teniendo en cuenta que su inviolabilidad no queda circunscrita constitucionalmente a las personas físicas, pues el artículo 18 de la Constitución no distingue entre personas físicas y jurídicas, y porque el Tribunal Constitucional la ha proclamado también respecto a las personas jurídicas con las adaptaciones necesarias, pues la intimidad, inherente a la dignidad humana, no desempeña el mismo papel respecto a las personas jurídicas. Algo similar ocurriría con los sistemas de inteligencia artificial personificados, en los que la protección del domicilio aparecería más vinculada a la tutela de la actividad empresarial y económica, como ocurre con los seres humanos, que con el desarrollo de la intimidad al amparo de miradas ajenas indiscretas. Al margen de ello, cabe apuntar que la aparición de una personalidad jurídica electrónica o algorítmica supondría la consolidación definitiva del domicilio como sede virtual o digital, desvinculada de la idea tradicional del domicilio como sede física donde se desarrolla la intimidad a salvo de toda mirada indiscreta.

- El honor y la propia imagen

Igualmente reconocidos para las personas jurídicas, con las debidas adaptaciones respecto a las personas físicas, también serían predicables respecto a las personas electrónicas en cuanto que en esos derechos existe no solo una concepción personalista, sino también una perspectiva objetiva equiparable a la buena reputación. Y es que, si el fundamento último de la atribución de estos derechos de corte personalista a las personas jurídicas se basa, tal y como ha reconocido el Tribunal Constitucional, en que el derecho de asociación es un derecho de rango constitucional (lo que implica el reconocimiento de la titularidad de aquellos derechos que sean necesarios para la consecución de los fines de aquellas personas jurídicas), esa misma argumentación será trasladable a las, eventualmente, reconocidas como personas algorítmicas o electrónicas por cuanto los derechos a la libre investigación y producción científica y técnica, fundamento de esa personificación (tal y como se ha expuesto antes), gozan del mismo rango constitucional, por lo que resulta lógico que se le lleguen a reconocer también todos aquellos derechos que sean necesarios y complementarios para la consecución de sus fines.

- La libertad de creación técnica y artística

Son dos ámbitos que pueden sufrir una profunda transformación como consecuencia de la proliferación de los sistemas de inteligencia artificial con independencia de que estos lleguen a ser reconocidos como sujetos de derecho. La clave para la atribución de la titularidad de derechos de propiedad industrial e intelectual radicarán, como es lógico, en la determinación de quién es el sujeto artífice de la actividad inventiva y de la creación de la obra.

- El matrimonio

Sin mencionar de manera directa el matrimonio, tanto el Proyecto de informe como la Resolución del Parlamento Europeo, se muestran precavidos frente a los posibles vínculos emocionales entre seres humanos y robots, y consideran que esos vínculos pueden generar graves consecuencias físicas y emocionales en detrimento de la dignidad humana, lo cual sería en especial preocupante en el caso de personas necesitadas de especial protección (niños, ancianos y discapacitados). Frente a esta postura restrictiva, algunos expertos en inteligencia artificial han aventurado que los matrimonios con robots serán una realidad en apenas tres décadas. En cualquier caso, el análisis de los aspectos jurídicos de las relaciones y los vínculos emocionales entre seres humanos y sistemas de inteligencia artificial obliga

a tener en cuenta también la posición jurídica del ser humano y recordar que, precisamente, la dignidad es fundamento del libre desarrollo de la personalidad (protegido por el artículo 10 de la Constitución de 1978, al igual que en otras Constituciones europeas, e implícito en la Convención Europea de Derechos Humanos), y una regulación prohibitiva de vínculos emocionales entre sistemas de inteligencia artificial y seres humanos supondría una restricción de los derechos de estos últimos. No puede ignorarse que la robótica está llamada a desempeñar un papel relevante en el ámbito de las relaciones personales, sobre todo en el sexual, tanto en general como para las personas con diversidad funcional, y más que una llamada al paternalismo restrictivo, lo que procede es la puesta en valor, una vez más, del principio de autonomía y del consentimiento informado, por supuesto, con todas las precauciones que sean necesarias en el caso de los discapacitados.

En cuanto a la fiscalidad de las rentas económicas generadas por estas personalidades electrónicas y su eventual contribución a la Seguridad Social, eran dos aspectos centrales en el Proyecto de informe, que incluso planteó la posibilidad de financiar con esos ingresos una renta básica universal; todo ello queda difuminado en la Resolución, que se limita a destacar

[...] la importancia que reviste la previsión de los cambios sociales, habida cuenta de los efectos que podrían tener el desarrollo y la implantación de la robótica y la inteligencia artificial; (y se) pide a la Comisión que analice los diferentes posibles escenarios y sus consecuencias para la viabilidad de los sistemas de seguridad social en los Estados miembros;

En cuanto a los requisitos formales para la atribución de la personalidad jurídica, el Proyecto y la Resolución destacan (en general y como requisito previo a la comercialización de robots) la conveniencia de un sistema global de registro de robots, proponiendo la creación de una Agencia de la Unión Europea para la robótica y la inteligencia artificial, a la que se atribuiría la gestión del sistema de registro y de las inscripciones.

En esta línea, la inscripción de un robot o de un sistema de inteligencia artificial (o de una categoría de estos, según prevé la Resolución) con la asignación de un código o denominación identificativa sería un requisito que culminaría el proceso constitutivo de esta nueva categoría de personalidad jurídica.

Sin embargo, aunque otros registros ya existentes (el registro mercantil) también cumplen ese papel constitutivo de la personalidad jurídica (de las so-

ciedades mercantil), no parece que su modo de llevanza, a cargo de un funcionario, que califica la legalidad de todos los actos que acceden al registro, pueda trasladarse a este hipotético registro europeo de robots, pues más allá de ser requisito constitutivo de la personalidad jurídica, la razón esencial que justifica la creación de este registro específico y global es la trazabilidad. Y esta no se construye con datos jurídicos o económicos de los que pueda tomarse razón mediante asientos en un registro tradicional previa calificación de su legalidad, sino que se trata de asegurar que todas las etapas de toma de decisión del robot puedan ser objeto de reconstrucción, de modo que siempre ha de ser posible justificar cualquier decisión tomada por el robot, lo que exige constancia en tiempo real de los «datos de todas las operaciones efectuadas por la máquina, incluidos, en su caso, los pasos lógicos que han conducido a la formulación de sus decisiones».

En cualquier caso, queda claro que la inscripción en ese futuro registro sería un requisito constitutivo de la personalidad jurídica algorítmica o electrónica, pero no sería el único. Así, la Resolución habla también de la creación de un fondo patrimonial para garantizar la responsabilidad por daños, y propone dotar a estos sistemas de una

[...] matrícula individual que figure en un registro específico de la Unión que asegure la asociación entre el robot y el fondo del que depende y que permita que cualquier persona que interactúe con el robot esté al corriente de la naturaleza del fondo, los límites de su responsabilidad en caso de daños materiales, los nombres y las funciones de los participantes y otros datos pertinentes.

Una vía a explorar a la hora de configurar este futuro registro específico global de la Unión Europea para los robots avanzados con garantía de trazabilidad (técnica y jurídica) y publicidad del fondo patrimonial del robot sería la posibilidad de acudir a la tecnología de las cadenas de bloques para el registro de la matrícula del sistema, fondo patrimonial y demás circunstancias susceptibles de inscripción en el registro, que quedaría así distribuido en multitud de nodos, papel que correspondería a cada uno de los robots registrados. Ese mismo sistema permitiría la tokenización de participaciones o cuotas en el fondo patrimonial del robot (su negociación en mercados) y representativas del derecho a participar en los beneficios que genere el robot y en el reparto del fondo a la extinción de la personalidad jurídica del sistema.

La normativa reguladora de esta materia debería ocuparse de la configuración jurídica del acto dotacional de ese fondo que constituiría el patrimonio inicial de la persona algorítmica o electrónica: a cargo del fabricante, del co-

ercializador o, en general, de cualquier agente interviniente en el proceso productivo del robot o sistema de inteligencia artificial; como acto único, o mediante dotaciones periódicas; disponibilidad del fondo y requisitos para la gestión y, en su caso, su rentabilización. En el caso de sistemas o robots que desarrollen actividades que reporten un beneficio económico, cabría la apuntada posibilidad de tokenización de las participaciones en el fondo con posibilidad de negociación en mercados de valores, y representativas del derecho a participar en los beneficios de aquellas y en el reparto del fondo a la extinción de la personalidad jurídica del sistema.

Otro requisito propuesto por la Resolución del Parlamento Europeo es el establecimiento de un régimen de seguro obligatorio. Este podría imponerse a cargo del fabricante como contrapartida a su liberación de responsabilidad, que pasaría a recaer sobre el nuevo sujeto de derecho y se entendería sin perjuicio de los que eventualmente pudiese contratar este en el ejercicio de sus actividades.

Como es lógico, todo el proceso constitutivo y dotacional de la persona algorítmica o electrónica debería ser de tramitación telemática y estar autogestionado por ella misma, en consonancia con su carácter autónomo. Por ello, la creación de esta nueva categoría de persona algorítmica debería ir acompañada de la implantación de entornos y sistemas por parte de las administraciones públicas, bancos, aseguradoras, fabricantes y demás agentes actuantes en este mercado, que permitan la gestión autónoma por *software* inteligente diseñado a tal efecto.

Siguiendo con los requisitos necesarios para la atribución de esta nueva categoría de personalidad jurídica, hay algunos que se relacionan de manera directa con el principio de transparencia al que se ha hecho referencia con anterioridad. En ese sentido, y siendo la consecuencia de esta personificación la liberación de responsabilidad del fabricante, su contrapartida debería ser el depósito del código fuente (de manera preferible en el mismo registro antes mencionado), la restricción del plazo de duración de las patentes y la flexibilidad en la concesión de licencias de explotación.

Bibliografía

- ASÍS ROIG, Rafael de (2013). *El marco ético de la robótica*. Madrid: Instituto de Derechos Humanos Bartolomé de las Casas, Departamento de Derecho Internacional, Eclesiástico y Filosofía del Derecho Universidad Carlos III de Madrid. https://e-archivo.uc3m.es/bitstream/handle/10016/19290/marco_asis_PTD_2013.pdf.
- BADIA, J. G.; DE PUELLES, P. G. y DE LECUONA, I. (2018). «Big data y seguridad de la información». *Atención primaria*, vol. 50, núm. 1, págs. 3-5. Disponible en: www.bioeticayderecho.ub.edu/es/editorial-big-data-y-seguridad-de-la-informacion.
- BARRIO, Andrés Moisés y GARCÍA-PRieto CUESTA, Juan (2018). *El Derecho de los robots*. 1.^a ed. Madrid: Wolters Kluwer.
- BAYERN, Shawn (2015). «The Implications of Modern Business-Entity Law for the Regulation of Autonomous Systems». *Stanford Technology Law Review* 93 FSU College of Law, Public Law Research Paper núm. 797. <https://ssrn.com/abstract=2758222>.
- BAYERN, Shawn; BURRI, Thomas; GRANT, Thomas D.; HÄUSERMANN, Daniel M.; MÖSLEIN, Florian y WILLIAMS, Richard (2017). «Company Law and Autonomous Systems: A Blueprint for Lawyers, Entrepreneurs, and Regulators». *Hastings Science and Technology Law Journal* 2, 135-162. <https://ssrn.com/abstract=2850514> or <http://dx.doi.org/10.2139/ssrn.2850514>.
- BERNAL CRESPO, Julia Sandra (2011). *Evolución Biológica de la Moral y el Derecho*. 1.^a ed. Bogotá: Universidad del Norte.
- BROSNAN, Sarah (2013). *Comportamientos relacionados con la justicia y la equidad en primates no humanos*. Trad. Bernardo Yáñez. *Ludus Vitalis*, vol. XXI, núm. 40.
- BUISÁN, L. y URRUTIA, A. S. (2011). *Intimidación, confidencialidad y protección de datos de salud: aportaciones del IV Seminario Internacional sobre la «Declaración Universal sobre Bioética y Derechos Humanos» de la Unesco*. Cizur Menor: Civitas.
- CALO, Ryan (2011). «Open Robotics». *Maryland Law Review*, vol. 70, núm. 3. <https://ssrn.com/abstract=1706293>.
- *Robots in American Law*. University of Washington - School of Law; Paper núm. 2016-4 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2737598.
- CASADO, M. (1998). «Ética, Derecho y deontología profesional». *DS: Derecho y salud*, vol. 6, núm. 1, págs. 30-35. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=2891725>.
- (2007). *Nuevos materiales de bioética y derecho*. México: Fontamara.

- (2009). *Sobre la dignidad y los principios. Análisis de la Declaración Universal sobre Bioética y Derechos Humanos de la Unesco*. Cizur Menor: Civitas.
- (2017). *De la solidaridad al mercado. El cuerpo humano y el comercio biotecnológico*. Barcelona: Edicions de la Universitat de Barcelona.
- CASADO, M. y LÓPEZ BARONI, M. J. (2018). *Manual de bioética laica (I). Cuestiones clave*. Barcelona: Edicions de la Universitat de Barcelona.
- CASTIELLA RODRÍGUEZ, José Javier (2003). *Instituciones de Derecho Privado*. Tomo I, volumen 2. 1.^a ed. Madrid: Consejo General del Notariado, Civitas Ediciones SL.
- CASTRO y BRAVO, Federico de (2008). *Derecho Civil de España*. Tomo II. 1.^a reed. Aranzadi.
- COMISIÓN EUROPEA (2018). *Inteligencia artificial, robótica y sistemas «autónomos»*. Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías. Traducción de Maria José Villalobos, miembro del OBD. Disponible en: www.bioeticayderecho.ub.edu/archivos/pdf/EGE_inteligencia-artificial.pdf.
- CUENCA GÓMEZ, Patricia (2012). «Sobre la inclusión de la discapacidad en la teoría de los derechos humanos». *Revista de Estudios Políticos* (nueva época), núm. 158, Madrid, octubre-diciembre 2012, págs. 103-137.
- DONEAU, Hugues (1846). *Hugonis Donelli jurisconsulti et antecessoris opera omnia 6 Commentariorum de jure civil. Tomus primus*. Ed. ad signum Cluis Florencia (consultado en books.google.es).
- Draft Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence*. European Commission Directorate-General for Communication (2018). Bruselas: European Commission.
- DE LECUONA, I. (2019). «Evaluación de los aspectos metodológicos, éticos, legales y sociales de proyectos de investigación en salud con datos masivos (Big Data)». *Gaceta Sanitaria*, vol. 32, núm. 6, págs. 576-578. Disponible en: <https://doi.org/10.1016/j.gaceta.2018.02.007>.
- ENGELHARDT, H. Tristram (1995). *The Foundations of Bioethics*. 2.^a ed. Oxford University Press.
- FERNÁNDEZ SESSAREGO, Carlos (2001). *¿Qué es ser persona para el Derecho?* Derecho Privado, libro en homenaje a Alberto J. Bueres, Buenos Aires: Editorial Hammurabi.
- FERRARA, FRANCESCO (2006). *Teoría de las Personas Jurídicas II*. Trad. de la 2.^a edición: Eduardo Ovejero y Mauri. Granada: Editorial Comares.
- FLORIDI, Luciano y SANDERS, J. W. (2004). *On the morality of artificial agents*. University of Oxford. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.16.722&rep=rep1&type=pdf>.
- GAGARIN, Michael (1989). *Early Greek Law*. University of California Press.
- GARCÍA MANRIQUE, R. (2018). *El cuerpo diseminado. Estatuto, uso y disposición de los biomateriales humanos*. Cizur Menor: Civitas / Thomson Reuters / Aranzadi.
- GÜNTHER, Jakobs (2005). *¿Derecho penal del enemigo? Un estudio acerca de los presupuestos de juridicidad*. Trad. Manuel Cancio Meliá. www.panoptica.org/seer/index.php/op/article/viewFile/Op_2.7_2007_197-213/280.

- HALL, Robert T. (2015). *Los derechos de los animales*. Madrid: Marcial Pons.
- HARARI, Yuval Noah (2016). *Homo Deus. Breve historia del mañana*. 3.^a ed. Barcelona: Editorial Penguin Random House.
- HARRIS, Marvin (1986). *Caníbales y Reyes*. 1.^a ed. Barcelona: Salvat Editores.
- HART, HERBERT L. A. (1994). *The concept of Law*. 2.^a ed. Editado por Joseph Raz y Penelope A. Bulloch. Clarendon Law Series.
- HERVÁS ARMERO, María Dolores (2010). *Derecho, Biotecnología y principios éticos. Contextos sociales y legales. Normativa y Bioética comparada: especial atención al caso de Japón*. Valencia: Universitat de Valencia. <http://roderic.uv.es/bitstream/handle/10550/23465/hervas.pdf?sequence=1&isAllowed=y>.
- HUSSON, Genevieve y VALBELLE, Dominique (1998). *Instituciones de Egipto: De los primeros faraones a los emperadores romanos*. Trad. Maribel García Sánchez. Madrid: Cátedra.
- IGLESIAS, Juan (1983). *Derecho Romano*. 8.^a ed. Barcelona: Ariel.
- KANT, Immanuel (1999). *Teoría y praxis*. El aleph.com.
- LLÀCER, M. R., CASADO, M., & BUISÁN, L. (2015). *Documento sobre bioética y Big Data de salud: explotación y comercialización de los datos de los usuarios de la sanidad pública*. Barcelona: Publicacions i Edicions de la Universitat de Barcelona. Disponible en: <http://hdl.handle.net/2445/104585>.
- LOCKE, John (2005). *Segundo tratado sobre el gobierno civil. Un ensayo concerniente al verdadero origen, alcance y finalidad del gobierno civil*. Trad. Claudio Amor y Pablo Stafforini. 1.^a ed. Argentina: Universidad Nacional de Quilmes.
- LÓPEZ BARONI, M. J. (2016). *El origen de la bioética como problema*. Barcelona: Edicions de la Universitat de Barcelona.
- LÓPEZ BARONI, M. J. (2018). *Las narrativas de la biotecnología*. Argumentos de Razón Técnica.
- LÓPEZ BARONI, M. J. (2019). «Las narrativas de la inteligencia artificial». *Revista de Bioética y Derecho*, núm. 46, págs. 5-28. Disponible en: <http://revistes.ub.edu/index.php/RBD/article/view/27280>.
- MARTÍN VALENTÍN, FRANCISCO J. y LARA PEINADO, Federico (1999). *Sobre la utilización y organización de las Fundaciones Funerarias durante el Imperio Antiguo Egipcio*. *GERION, Revista de Historia Antigua*, vol. 17. <https://dialnet.unirioja.es/servlet/articulo?codigo=101327>.
- MATTHIAS, Andreas (2004). *The responsibility gap: Ascribing responsibility for the actions of learning. Ethics and Information Technology* 6: 175-183. Países Bajos: Kluwer Academic Publishers. <https://link.springer.com/article/10.1007/s10676-004-3422-1>.
- MEGÍAS QUIRÓS, José Justo (2016). «El Derecho y los derechos en la antigua Grecia». *Revista Dikaioσύne*, núm. 31. Universidad de los Andes. www.researchgate.net/publication/319988284_El_Derecho_y_los_derechos_en_la_Antigua_Grecia.
- MONASTERIO ASTOBIZA, Aníbal (2019). «Ética para máquinas: Similitudes y diferencias entre la moral artificial y la moral humana». *Ilemata, Revista Internacional de éticas Aplicadas*. Año 11, núm. 30, 129-147.

- MOSTERÍN, Jesús (2015). *Los derechos de los animales*. Madrid: Marcial Pons.
- MOOR, James H. (2006). *The Nature, Importance, and Difficulty of Machine Ethics Intelligent Systems* (IEEE). www.researchgate.net/publication/220629129_The_Nature_Importance_and_Difficulty_of_Machine_Ethics.
- NEVEJANS, Natalie (2016). *Study for Jury Committee European Civil Law Rules In Robotics Policy*. Bruselas: Departamento C: Citizens Rights and Constitutional Affairs European Parliament www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU%282016%29571379_EN.pdf.
- NUSSBAUM, Martha C. (2007). *Las fronteras de la justicia. Consideraciones sobre la exclusión*. Barcelona: Paidós Ibérica.
- OHELING DE LOS REYES, Alberto (2011). «El Concepto constitucional de dignidad de la persona. Forma de comprensión y modelos predominantes de recepción en la Europa continental». *Revista española de derecho constitucional*. núm. 91, págs. 135-178.
- O'NEIL, Cathy (2017). «Armas de destrucción matemática. Cómo el big data aumenta la desigualdad y la democracia». Trad. Violeta Arranz de la Torre. Madrid: Capitan Swing Libros.
- PALMERINI, Erika (2017). «Robótica y derecho: sugerencias, confluencias, evoluciones en el marco de una investigación europea». *Revista de Derecho Privado*, Universidad Externado de Colombia, núm. 32. <https://doi.org/10.18601/01234366.n32.03>.
- POLO SANTILLÁN, Miguel Ángel (2004). «Bioética y persona en Peter Singer». *Escritura y Pensamiento*. Año VII, núm. 14.
- Proyecto de informe de 31 de mayo de 2016 con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica* (2015/2103(INL).
- RAWLS, John (1995). *Teoría de la justicia*. Trad. de María Dolores González. Reimpresión 2006. México: Fondo de Cultura Económica.
- REGIS, Fatima; TIMPONI, Raquel y MAIA, Alessandra (2011). *Cognición integrada, encadenada y distribuida: breve discusión de los modelos cognitivos en la cibercultura*. www.thefreelibrary.com/Cognicion+integrada%2c+encadenada+y+distribuida%3a+breve+discusion+de+los...-a0339733357.
- 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).
- RODRÍGUEZ CORONEL, Martha M. (2012). «A vueltas con la agencia moral. Una perspectiva crítica». *Quaderns de filosofia i ciència*, 42, págs. 127-138.
- ROTH, Martha T. (1995). *Law Collections from Mesopotamia and Asia Minor*. Boston: Society of Biblical Literature.
- RUSSEL, Stuart (2017). *El próximo paso: la vida exponencial*. Colección Open Mind BBVA. www.bbvaopenmind.com/wp-content/uploads/2017/01/BBVA-Open-Mind-libro-El-proximo-paso-vida-exponencialr.pdf.
- SÁNCHEZ ZORRILLA, Manuel y Zavaleta Chimbor, David (2011). «El Derecho en una sociedad ágrafa: investigación jurídico-epistemológica sobre el Derecho y el De-

- recho consuetudinario en la sociedad inca». *Revista Telemática de Filosofía del Derecho*, núm. 14.
- SINGER, Peter (1995). *Ética Práctica*. Cambridge University Press.
- SUPIOT, Alain (2007). *Homo Juridicus: Ensayo sobre la función antropológica del derecho*. Madrid: Siglo XXI de España Editores.
- VAN WYNSBERGHE, Aimee y Robbins, SCOTT (2018). *Critiquing the Reasons for Making Artificial Moral Agents*. <https://link.springer.com/article/10.1007%2Fs11948-018-0030-8>.
- WILKER, Daniel (2017). *Paternalismo en la época del mejoramiento cognitivo: ¿las libertades civiles presuponen aproximadamente las mismas habilidades mentales? Mejoramiento humano*. Trad. Inés Ramia y Alicia Jiménez. 1.^a ed. Zaragoza: Teell editorial.
- WILLIAMS, Andrew P. y SCHARRE, Paul D. (2015). *Autonomous Systems Issues for Defence Policymakers*. Países Bajos: NATO Communications and Information Agency. www.researchgate.net/publication/282338125_Autonomous_Systems_Issues_for_Defence_Policymakers.

Del mismo modo que el vapor, la electricidad y la computación fueron los pilares de las tres revoluciones industriales que han modelado nuestra sociedad, la cuarta se erige sobre la robótica y la inteligencia artificial.

Los desafíos que plantean estos sistemas inciden en el campo ético y jurídico, entre otras razones porque la creciente autonomía de la inteligencia artificial puede romper los esquemas que tradicionalmente determinan la responsabilidad por daños. En la búsqueda de posibles soluciones al problema, la Comisión de Asuntos Jurídicos del Parlamento Europeo ha considerado la posibilidad de reconocer personalidad jurídica a los sistemas de inteligencia artificial más avanzados.

Esta obra se suma al debate originado por dicha propuesta enlazando una eventual personalidad jurídica electrónica o algorítmica con la evolución histórica del concepto de persona como sujeto de derechos, y proponiendo criterios para su regulación sobre la base de cuatro principios de profunda raigambre bioética: beneficencia, no maleficencia, autonomía y justicia.