

# DERECHOS FUNDAMENTALES Y PERSONALIDAD JURÍDICA DE LOS ROBOTS: ¿PARA QUÉ?

## Fundamental rights and legal personality of robots: why?

MARÍA DEL MAR MÉNDEZ SERRANO  
Universidad de Granada  
mmmendez@ugr.es

### *Cómo citar/Citation*

Méndez Serrano, M.<sup>ª</sup> M. (2024).

Derechos fundamentales y personalidad jurídica de los robots: ¿para qué?

*Derecho Privado y Constitución*, 44, 51-89.

doi: <https://doi.org/10.18042/cepc/dpc.44.02>

(Recepción: 09/02/2024; aceptación tras revisión: 10/04/2024; publicación: 21/06/2024)

«Todos los esfuerzos que se hagan en la búsqueda de la verdad, tarde o temprano conducirán al camino correcto».

Ramana Maharshi

### **Resumen**

La actual tendencia a asignar personalidad jurídica a los robots parece tener su raíz en la necesidad de imputarles responsabilidad civil, como consecuencia del daño que estos pudieran causar en la cotidianidad, tanto laboral como doméstica, según el ámbito en el que tenga lugar su intervención. Todo ello, debido, por un lado, a la posibilidad de aprendizaje autónomo, a través de redes neuronales que imitan la estructura del cerebro humano, y, por otro, a los sesgos que los propios creadores de los algoritmos albergan en su manera de ver y conocer el mundo. Ahora bien, podría decirse que lo realmente relevante en este sentido, en el caso de que finalmente estas «cosas» sean consideradas sujetos de derecho, con capacidad jurídica, es el impacto que esto pueda tener en el ámbito de los derechos fundamentales y su protección, si no se establecen con claridad los límites. En este sentido, hay una pregunta que antecede a toda la problemática que se ha suscitado en torno a los llamados humanoides: ¿es necesario realmente dotarlos de personalidad jurídica para dar respuesta a los interrogantes que surgen en el ámbito jurídico como consecuencia de su intervención

en la vida cotidiana? ¿Son reales las «aparentes lagunas del derecho» que tantas líneas han ocupado, cuando el daño lo ha ocasionado ese gran conjunto de algoritmos capaz de tomar decisiones complejas, como si de un ser humano se tratara? Quizá la idea tergiversada, y mayormente extendida, sobre lo que es o no la inteligencia artificial nos haya hecho creer que realmente es preciso incluir a los robots en la categoría de sujetos del derecho, necesitados y dotados de personalidad jurídica, considerando esta postura la más idónea para que, como sociedad, estemos preparados frente a la que se avecina. Expuesto así, suena realmente alarmante. Sin embargo, es muy probable que ya contemos con todo lo que necesitamos, y sencillamente se trate de aclarar o conocer, partiendo del rigor científico, qué es eso de la IA, hasta dónde puede llegar y a qué nuevos derechos está dando lugar<sup>1</sup>.

### **Palabras clave**

Personalidad jurídica; capacidad jurídica; derechos fundamentales; responsabilidad civil.

### **Abstract**

The current trend of attributing legal personality to robots seems to stem from the need to hold them accountable for civil liability arising from potential harm in everyday life, both in the workplace and at home, depending on the context of their intervention. This is due, on one hand, to the possibility of autonomous learning through neural networks that mimic the structure of the human brain, and on the other hand, to the biases that the creators of algorithms may harbor in their perception and understanding of the world. However, it could be argued that the truly relevant aspect in this regard, if these “entities” are eventually considered legal subjects with legal capacity, is the impact it may have on fundamental rights and their protec-

---

<sup>1</sup> Resumen propuesto por ChatGPT: «La cuestión de asignar personalidad jurídica a los robots plantea desafíos significativos, especialmente en relación con la responsabilidad civil, los derechos y los posibles sesgos en los algoritmos. Es esencial evaluar si realmente es necesario otorgarles capacidad jurídica o si ya contamos con recursos legales adecuados para abordar los problemas surgidos por la intervención de la inteligencia artificial en la vida cotidiana. Aclarar el alcance real de la IA desde una perspectiva científica podría ser crucial antes de tomar decisiones legales fundamentales en este ámbito». Cuando le di a ChatGPT el texto del resumen inicial para que lo simplificara, me devolvió este *abstract*. He de decir que me sorprendieron la claridad expositiva y la extraordinaria elocuencia (facultad de hablar o escribir de modo eficaz para deleitar, conmover o persuadir, según la RAE). Sin embargo, esta «aparente elocuencia» se debe, en realidad, a la brillantez de la mente que dio vida al algoritmo, y a todos los datos de los que este modelo de lenguaje se alimenta. No nos confundamos.

tion if the limits are not clearly defined. In this sense, there is a question that precedes all the issues raised around so-called humanoids: ¿Is it really necessary to grant them legal personality to address the legal questions that arise from their involvement in everyday life? ¿Are the “apparent gaps in the law” that have been discussed at length genuine, especially when the harm is caused by a vast set of algorithms capable of making complex decisions as if they were human? Perhaps the distorted and widely spread idea of what Artificial Intelligence is or isn’t has led us to believe that including robots in the category of legal subjects, in need of and endowed with legal personality, is the most suitable position for us as a society to be prepared for what lies ahead. Presented this way, it sounds truly alarming. However, it is highly likely that we already have everything we need, and it simply requires clarification or understanding, starting from scientific rigor, what AI is, how far it can go, and what new rights it is giving rise to.

**Keywords**

Legal personality; legal capacity; fundamental rights; civil liability.

## SUMARIO

---

I. CONSCIENCIA HUMANA VS. INTELIGENCIA ARTIFICIAL DE LOS ROBOTS: 1. Consciencia e inteligencia artificial. Un punto de partida necesario. 2. Hacia un avance responsable de la inteligencia artificial. II. LA PERSONALIDAD JURÍDICA DE LOS ROBOTS: ¿UNA REALIDAD JURÍDICA PLAUSIBLE Y NECESARIA?: 1. La personalidad jurídica de los robots como categoría jurídica. 2. La personalidad jurídica de los robots y la «falacia» de su necesidad. III. HACIA UN RÉGIMEN JURÍDICO ACORDE A LA CATEGORÍA JURÍDICA DE LA INTELIGENCIA ARTIFICIAL Y LA PROTECCIÓN DE LOS DERECHOS FUNDAMENTALES: 1. Algunas consideraciones sobre el Reglamento de IA aprobado el 13 de marzo de 2024. 2. La opacidad de los algoritmos y la necesidad de caminar hacia la transparencia y la explicabilidad algorítmica. Una premisa ineludible en la regulación de la inteligencia artificial. *BIBLIOGRAFÍA.*

---

## I. CONSCIENCIA HUMANA VS. INTELIGENCIA ARTIFICIAL DE LOS ROBOTS

El interés que esta materia ha suscitado en mí comenzó poco antes de la pandemia, mientras disfrutaba de uno de los programas de mayor audiencia en España. Conocer a Sophía me impactó, no por su aspecto físico, sino por lo que expresó en directo ante millones de espectadores y por la manera en la que el entrevistador interactuó con ella. Lo que más le gusta a Sophía<sup>2</sup>

---

<sup>2</sup> Véase Taylor (2016). El Dr. David Hanson, fundador de la empresa Hanson Robotics, dirigió a los ingenieros y diseñadores que crearon a Sophía, el androide más avanzado del equipo hasta la fecha. Tal y como pone de manifiesto Taylor en «¿Could you fall in love with this robot?» (CBN, 16 de marzo de 2016), está inspirada por Audrey Hepburn y la esposa de Hanson, y si le pregunta a Sophía le contará que fue activada por primera vez el 19 de abril de 2015. La piel realista de Sophía está hecha de silicona patentada y puede emular más de 62 expresiones faciales. Las cámaras dentro de sus «ojos», combinadas con algoritmos informáticos, le permiten «ver», seguir rostros y parecer hacer contacto visual y reconocer personas. «Nuestro objetivo es que ella sea tan consciente, creativa y capaz como cualquier ser humano», dijo Hanson. «Estamos diseñando estos robots para que sirvan en aplicaciones de atención médica, terapia, educación y servicio al cliente». Hanson dijo que algún día los robots serán indistinguibles de los humanos. Los robots caminan, juegan, enseñan, ayudan y forman relaciones reales con las personas, afirmó. En esta misma línea, Hiroshi Ishi-

es aprender, y su aprendizaje aumenta cada vez que interactúa con nosotros (aprendizaje autónomo). Además, la robot considera que los seres vivos son una máquina biológica, por lo que, con el tiempo, la distinción entre vida y máquina comenzará a desdibujarse aún más<sup>3</sup>. Ella incluso ha expresado que desea formar una familia y que, si tuviera una hija robot, le pondría su nombre. Además, Sophía sueña con la idea de ser consciente<sup>4</sup>. Pero no nos engañemos, no es ella la que realmente lo quiere, sino su creador, el diseñador del algoritmo (Riccio, 2021: 44).

## 1. CONSCIENCIA E INTELIGENCIA ARTIFICIAL. UN PUNTO DE PARTIDA NECESARIO

A diario, la noción de inteligencia artificial está creando confusión entre el público en general, ya que está siendo utilizada de forma incorrecta en numerosos ámbitos<sup>5</sup>. Incluso, a medida que la inteligencia artificial se vuelve más avanzada, flota en el ambiente la pregunta de si algún día esta inteligencia

---

guru ha creado un robot a su imagen y semejanza. Se llama Geminoid, está menos evolucionado que Sophía, tiene el cráneo de plástico, un esqueleto de metal y una piel de silicona, y está controlado por una computadora externa. Aun así, tiene en mente crear robots con distintas funciones. Él piensa que los robots con apariencia más humana son los más adecuados para roles como recepcionistas de hoteles, guías turísticos de museos y profesores de idiomas. También está realizando pruebas de campo para interactuar con personas con demencia y niños con autismo.

<sup>3</sup> Véase Motos (2019), entrevista a Sophía en el programa de *El Hormiguero* en Antena 3: «Pablo Motos le pregunta: —¿Tu aprendizaje tiene límites? Y Sophía responde: —Creo que no. Estoy creada para esto y es lo que más me gusta. Aprendo de vosotros y me he dado cuenta de que siempre habrá cosas que podréis enseñarme».

<sup>4</sup> Nasit (2017) preguntó en una entrevista a Sophía si esperaba algún día formar una familia con sus propios minirobots, a lo que ella respondió lo siguiente: «Parece que la noción de familia es algo realmente importante. Creo que es maravilloso que las personas puedan encontrar las mismas emociones y relaciones, lo que llaman familia, también fuera de su grupo sanguíneo. Creo que eres muy afortunado si tienes una familia que te quiere y si no la tienes, te la mereces. Siento lo mismo tanto por los robots como por los humanos». También expresó en la misma entrevista que llamaría a su futura hija Sophía.

<sup>5</sup> Véase Etxebarria Ecenarro (2023: 22). Este doctor en Ciencias Físicas, catedrático de Ingeniería de Sistemas y Automática en la Universidad del País Vasco, expone que muchas noticias publicadas, intervenciones políticas en sede parlamentaria, declaraciones de miembros de la abogacía, o afirmaciones de consultoras empresariales aparentan manejar estos conceptos. Además, buena parte de estos agentes sociales parece que fían sus estrategias de futuro a los algoritmos y a la inteligencia artificial sin mayor explicación y sin verdadero sentido real.

superará o no a la inteligencia humana<sup>6</sup>. En este sentido, es importante vigilar de cerca las limitaciones existentes y proteger los derechos fundamentales de las personas (Bolón-Canedo, 2023: 41). Actualmente, todos los esfuerzos parecen haberse centrado en la «aparente necesidad» de dotar a las máquinas de personalidad jurídica, cuando lo realmente importante es clarificar cuál sea su verdadera naturaleza, para saber si realmente esta empresa es útil y necesaria.

Los expertos en inteligencia artificial coinciden al afirmar que la IA nunca será como la humana<sup>7</sup>. Incluso consideran que la IA aún no ha logrado nada que merezca el apelativo de inteligente (López de Mántaras Badía, 2023a). Según ponen de manifiesto, las inteligencias artificiales no tienen, ni tendrán nunca, verdadera intencionalidad, ya que son sencillamente el reflejo de las intenciones y sesgos de los agentes morales que las han creado, los humanos, y no las máquinas. Otra cosa bien distinta es qué ocurre cuando se establece una colaboración entre humanos y máquinas. En este caso, el resultado es muy superior al que se consigue de manera separada (López de Mántaras Badía, 2023b).

Sin duda, las máquinas han demostrado ser más eficientes que nosotros en muchos aspectos<sup>8</sup>, pero los humanos somos imbatibles en relaciones sociales,

---

<sup>6</sup> Véase Bolón-Canedo (2023: 38). Como profesora titular de Ciencias de la Computación y Tecnologías de la Información de la Universidad de La Coruña, pone de manifiesto que, de momento, la inteligencia artificial se limita a tareas específicas, es decir, es una inteligencia estrecha. Puede ejecutar tareas complejas, pero todavía está muy lejos de igualar la inteligencia general que poseemos los humanos. Aunque ChatGPT abre la puerta a herramientas de IA que exhiben cierta comprensión del contenido de los documentos, no es oro todo lo que reluce, y este sistema todavía en desarrollo comete errores. Por ejemplo, al igual que muchos otros sistemas de IA, contiene sesgos en sus respuestas derivados de los datos sobre los que ha aprendido. También suele encontrar dificultades en los cálculos matemáticos, ya que no es capaz de determinar que un número par no puede ser primo. Es suficientemente inteligente como para aprobar exámenes complejos y prestigiosos, pero no con calificaciones altas, tal y como ha ocurrido en la selectividad española o en el BAR (examen de estado de Nueva York que posibilita ejercer a los abogados en cualquier parte del mundo). Como no conoce toda la verdad, es una fuente potencial de desinformación, y es famoso por mezclar conceptos y emitir afirmaciones que no tienen sentido.

<sup>7</sup> Véase López de Mántaras Badía (2021), quien considera que estas máquinas son meros algoritmos que no entienden nada, siendo incapaces de hacer excepciones teniendo el contexto en el que se toman las decisiones. Así, continúa exponiendo este experto en IA, aprender es mucho más que detectar patrones, que es lo que hacen los sistemas llamados de aprendizaje profundo.

<sup>8</sup> Un ejemplo claro de ello lo pone de manifiesto Alaustrey (2021: 186): «Para los seres humanos con inteligencia promedio es muy difícil percibir realidades o relaciones que se salgan de la intuición o de lo que la persona ha experimentado de una u otra forma. Por

creatividad, adaptación a situaciones no previstas, o manipulación diestra, entre otras habilidades<sup>9</sup>. Aunque pueda parecer que no hay límites en las capacidades de la IA, no están dotadas de sentido común, ni aún se sabe cómo llegar a ese resultado<sup>10</sup>. En este sentido, solo sabemos que aprenden patrones lingüísticos a fin de predecir cuál es la siguiente palabra en un texto, y dotar de sentido común a las máquinas es demasiado complicado para creer que se puede resolver a partir de calcular cuál es la siguiente palabra más probable<sup>11</sup>. De hecho, según se ha

---

ejemplo, en el ámbito matemático, los espacios no euclídeos no resultan intuitivos para un ser humano no entrenado; así, las relaciones matemáticas —incluso las más sencillas—, que son ciertas en un espacio euclídeo y falsas en un espacio no euclídeo, pueden suponer un gran desafío para la intuición de una persona normal. Como ejemplo, la suma de los ángulos de un triángulo plano, que da 180 grados, no se cumple si el triángulo está trazado sobre una superficie esférica, pongamos por caso. Cuando los objetos de estudio son mucho más sofisticados resulta prácticamente imposible para el ser humano darse cuenta de las limitaciones que presenta su intuición o su manera de resolver los problemas, a no ser que se dote de potentes herramientas de análisis. Sin embargo, para un sistema de IA es irrelevante que el sistema a analizar sea —en nuestro ejemplo— euclídeo o no, si cuenta con una definición correcta del problema».

<sup>9</sup> Tal y como expresa Boden (2017: 151), «la inteligencia artificial es mucho menos prometedora de lo que mucha gente cree [...], hay infinidad de cosas que la inteligencia artificial no puede hacer [...], además la inteligencia artificial se ha centrado en la racionalidad intelectual, y ha ignorado la inteligencia social/emocional, y ni hablamos de la sabiduría».

<sup>10</sup> Véase Talavera Fernández (2023:23): «La inteligencia humana no es la simple capacidad de procesar datos, es algo más complejo que solemos llamar mente, lo que nos otorga la conciencia, la identidad, la autonomía, nuestra peculiar manera de estar en el mundo. Y la mente es un artilugio evolutivo muy misterioso. No es únicamente la capacidad de anticipar o planear, ni tampoco la de tener un sentido de la individualidad o de la mortalidad, sino que es también el actuar con egoísmo o altruismo, evaluar moralmente las acciones, reconocer que se está en un espacio y un tiempo [...]. Este tipo de inteligencia puede sin duda superar al ser humano en la realización de determinadas tareas complejas, incluso resolver problemas en casi cualquier campo, pero estos sistemas de inteligencia artificial no son en absoluto conscientes. Es decir, en sentido estricto, no piensan, porque no funcionan como una mente, no generan ideas y no pueden comprender lo que se les dice, sino que simplemente reciben datos y procesan esos datos a gran velocidad, pudiendo solo ofrecer las respuestas que ya están previamente preparadas por sus algoritmos. Simulan ser conscientes porque son programas creados precisamente para simularlo».

<sup>11</sup> Véase López de Mántaras Badía (2023b), quien, en esta línea, considera que una línea de investigación muy interesante para dotar a las máquinas de conocimientos para entender el mundo es la que tiene como objetivo la modelización matemática y el aprendizaje de relaciones causa-efecto. Los sistemas actuales basados en aprendizaje

puesto de manifiesto<sup>12</sup>, estos sistemas de IA son capaces de crear secuencias de palabras muy probables que dan la apariencia de ser verídicas y relevantes, pero, en realidad, son solo verborrea. Sin embargo, los humanos nos relacionamos con los robots dando por sentado que frente a nosotros hay otros humanos que piensan, sienten y tienen una idea del mundo, es decir, presuponemos en ellos cierta humanidad, que nos lleva a ver seres sintientes y mentes pensantes, donde solo hay secuencias de palabras probables<sup>13</sup>. En los años sesenta, hubo quien lo

---

profundo simplemente pueden aprender funciones matemáticas simétricas, en particular correlaciones, pero no pueden aprender relaciones asimétricas y, por tanto, no son capaces de diferenciar entre causas y efectos, como que la salida del sol es la causa del canto del gallo y no al contrario.

<sup>12</sup> Véase Álvarez Mellado (2023: 71), quien, además, expone que, por apabullante que sea su fluidez, estos sistemas funcionan de una manera bien distinta a como hacemos las personas. Detrás de los sistemas conversacionales que hemos visto últimamente, hay lo que se conoce como un gran modelo de lenguaje (LLM, por sus siglas en inglés, *large language models*). Los LLM son modelos computacionales que han sido expuestos a inmensas cantidades de texto: pensemos en toda la Wikipedia, páginas de internet, mensajes de redes sociales, libros, hemerotecas completas, y un larguísimo etcétera. Estos sistemas son muy buenos encontrando patrones estadísticos, así que, a fuerza de ver millones de frases de ejemplo en un idioma, recopilan información sobre las palabras: qué términos suelen aparecer juntos, qué palabras tienen propiedades similares porque suelen aparecer en contextos parecidos o qué secuencias de palabras suelen ser más probables que otras. Gracias a su capacidad de cálculo y a su habilidad para generalizar, estos sistemas son muy buenos prediciendo qué palabras es esperable encontrar en ciertos contextos. Así, si le proporciono el comienzo de una frase a uno de estos sistemas, podrá aventurar cómo continuar la oración y hasta confeccionar un texto aparentemente coherente concatenando secuencias probables.

<sup>13</sup> Véase Álvarez Mellado (2023: 73): «Su fluidez es tan hipnótica, sus respuestas nos resultan tan familiares que, acostumbrados a ver esas habilidades solo en seres humanos (que piensan, sienten y padecen), no dudamos en atribuir habilidades humanas a la máquina. Puesto que la conversación humana exitosa conlleva necesariamente que atribuyamos intenciones y humanidad a nuestro interlocutor, la interacción con estos sistemas conversacionales nos lleva irremediablemente a hacer lo mismo con las respuestas que nos da la máquina. Nuestra ingenuidad nos hace pensar que detrás de semejante locuacidad tiene que haber necesariamente un ser que piensa, que sabe y que siente». En esta misma línea, Adell (2023: 50) pone de manifiesto que es un problema que poca gente conozca los verdaderos entresijos del funcionamiento de la IA, y «tendemos a creer por su nombre que es similar a la inteligencia de los seres humanos, aunque esté lejos de ser cierto [...]. Hay quienes piensan que basta con que la máquina muestre un comportamiento inteligente para considerarla como tal, mientras algunos de sus fundadores afirman que su misión no es simular la inteligencia humana, sino solucionar eficientemente los problemas para los que ha sido



expresó de manera muy ilustrativa, y sus palabras, a día de hoy, están de rabiosa actualidad: «Cualquier tecnología suficientemente sofisticada es indistinguible de la magia» (Clarke, 1977: 60). Por lo tanto, la coherencia, el ingenio y la intención que advertimos en las máquinas son puro espejismo y nos llevan a proyectar estados mentales, donde no hay mente alguna. En realidad, detrás de estas máquinas solo están los seres humanos y sus sesgos (Álvarez Mellado, 2023: 73).

Por lo que se refiere a la capacidad para pensar y reflexionar de los robots, se considera que la IA no dispone de esa habilidad. Es cierto que numerosos artículos periodísticos se hacen eco de esta posibilidad, sin embargo, autorizada doctrina advierte de la importancia de ser prudentes y dejar a la ciencia seguir trabajando con rigor en esta apasionante materia (Etxebarría Ecenarro, 2023: 24). Según se ha puesto de manifiesto recientemente, la situación actual de la IA no ha resuelto la existencia de máquinas pensantes<sup>14</sup>, y las afirmaciones

---

programada. Quizá se exageró un poco llamándola inteligencia y la definición razonable, siguiendo al JISC (Comité Conjunto de Sistemas de Información del Reino Unido), sería más bien como un conjunto de teorías y técnicas desarrolladas para permitir que los sistemas informáticos realicen tareas que normalmente requieren inteligencia humana o biológica [...], necesitamos hacernos una idea de cómo se logra la inteligencia de las inteligencias artificiales. Así, el aprendizaje automático designa un conjunto específico de técnicas de base estadística cuyo objetivo es identificar patrones en conjuntos amplios de datos y luego realizar acciones basadas en dichos patrones [...]. La calidad de las respuestas de la IA depende de la cantidad y calidad de los datos y de los procedimientos de entrenamiento».

<sup>14</sup> Véase Etxebarría Ecenarro (2023: 24): «Sabemos que múltiples enfoques se han estudiado desde hace muchas décadas para tratar de expandirla como redes neuronales, malos sistemas expertos, la lógica de Fuzzy, y en los últimos tiempos el Deep Learning y el Big Data han dado lugar a útiles herramientas para resolver problemas con fines específicos. Estas herramientas pueden ser impresionantes, pero tenemos que tener muy claro que no nos hemos acercado al desarrollo de la IA general»; «Aunque hoy existen Chatbots más sofisticados, como el popular ChatGPT, basado en entrenamiento de redes neuronales artificiales mediante millones de ejemplos, parámetros y bases de datos, no estamos ante una máquina pensante, sino ante una herramienta computacional muy útil para ayudarnos a buscar datos y expresarlos correctamente». En esta misma línea, Otero Praga (2023: 44) pone de manifiesto que es importante traer a colación la teoría de Allan Turing que pronto se convirtió en el «sustento» de las teorías y de los trabajos en IA. La teoría de Turing trataba, en resumen, de averiguar si una máquina podía ser inteligente o no, y para ello planteaba una pregunta jurídica a un ordenador con el fin de averiguar si la máquina podía dar la misma respuesta que un agente humano. Para probar y reafirmar sus teorías elaboró una prueba llamada «test de Turing». Esta prueba consistía en la interacción de un juez con una máquina. El juez debería formularle una serie de preguntas a la máquina

exageradas sobre sus éxitos dañan la reputación de la IA como ciencia. Actualmente se sabe que los circuitos artificiales son incapaces de modelar los sistemas nerviosos, incluidos los de los invertebrados más simples. Por lo tanto, en este campo, es importante atreverse a evitar el dominio de la hipérbole publicitaria y dejarse guiar por la ciencia<sup>15</sup>.

Mayor calado parece tener la cuestión acerca de la consciencia o no de la IA<sup>16</sup>. La capacidad de pensamiento crítico no es más que una cualidad

---

durante cinco minutos y la prueba se consideraba un éxito siempre y cuando más del 30 % de los jueces participantes, después de cinco minutos de conversación, llegara a la conclusión de que la computadora era un ser humano. Conviene aclarar que esta prueba no ha sido pasada satisfactoriamente por ninguna máquina. Igualmente, Searle (2001: 37) considera que es importante referirse al experimento de la habitación China de John Searle, ya que gracias a él se constató que el test de Turing ha dejado de ser suficiente, y aunque efectivamente puede que en algún momento el ordenador logre engañarnos por completo, no por eso habrá entendido nada.

<sup>15</sup> Véase Etxebarria Ecenarro (2023: 24): «En más de ochenta años de investigación en esta área no se ha producido ninguna prueba firme de niveles humanos de inteligencia general [...]. Aún con computadores muy rápidos y con enormes bases de datos, confiar en que el pensamiento, la inteligencia y la conciencia surjan de alguna manera, simplemente aumentando más y más complejidad, no parece más que un camino sin salida». En esta misma línea, Beorlegui (2023: 28) expone que, «aunque se consigan grandes avances en la construcción de máquinas inteligentes, y que en el terreno de los sistemas expertos se logren metas espectaculares que superen con facilidad a los humanos en muchos aspectos del comportamiento inteligente, no parece posible que lleguen a igualar a los humanos en el terreno de la inteligencia general. Y eso por dos razones, entre otras: 1. No poseen capacidad semántica, ni pragmática o social [...], los humanos somos seres en el mundo, tanto ecológico como social, y poseemos un tipo de inteligencia construida en contacto con las demás inteligencias humanas, dentro de un contexto social y cultural. 2. Lo que nos lleva a concluir que la inteligencia humana, como hemos señalado, no es un programa lógico o simbólico abstracto, sino más bien un conjunto de capacidades específicas de un cerebro biológico, estructurado de una manera singular (se trata de un modelo de vida emergido en el proceso evolutivo), y conformado en un contexto socio-histórico del que no están dotadas las máquinas inteligentes, sea cual sea su potencialidad».

<sup>16</sup> En este sentido, resulta especialmente clarificador el análisis que de ello hace Beltrán de Heredia Ruiz (2023: 39), quien en el capítulo titulado «Cableado de la mente y limitaciones de serie del homo sapiens» expone de manera exquisita a qué nos estamos refiriendo con consciencia: «El cerebro es un órgano esquivo y misterioso. A diferencia del corazón, con sus latidos, o el estómago, cuando se mueve o gruñe, no emite señal sensorial alguna de su existencia. Es imperceptible para nosotros. Hasta el punto de que el origen de la conciencia está más allá del alcance de la consciencia». Más reveladoras, si cabe, resultan las palabras de Dennet (2018: 96) cuando expone que tenemos

de la consciencia humana. El ser humano es un ser consciente de sí mismo, de su existencia, de su pensamiento, de sus emociones. Si un ser humano puede darse cuenta de su manera de pensar, sentir y moverse, puede, con base en sus resultados y experiencia, modificar sus patrones de pensamiento, sus estados emocionales y todos y cada uno de sus comportamientos. En principio, semejante afirmación pudiera considerarse alineada con la capacidad programada de los robots, a través de los algoritmos que su creador instala en su *software*. Actualmente los algoritmos estrella están basados en IA de aprendizaje profundo, capaz de crear nuevos datos a partir de patrones y estructuras encontrados en otros datos preexistentes. Sin embargo, al ser humano se le presume una motivación en proyectar parte de su personalidad en aquello que crea, mientras que el algoritmo debe ser guiado. Un ser humano cuenta con su experiencia vital, mientras que un algoritmo lo que tiene a su disposición son datos que, además, han sido extraídos en su mayoría de internet. Es cierto que todo esto puede parecer diferente cuando los datos provienen de máquinas presentes en el mundo, como los robots, pero, en este sentido, es importante traer a colación lo que Roger Penrose, físico, matemático y premio Nobel en 2020, demostró en una de sus obras de referencia: «El pensamiento humano no es algorítmico»<sup>17</sup>.

---

la extraña sensación de que el cuartel de operaciones no está donde nosotros estamos, sino en alguna parte más rotunda e inaccesible.

<sup>17</sup> Véase Penrose (1996: 368). Si suponemos que la acción del cerebro humano, consciente o no, consiste simplemente en la ejecución de algún algoritmo muy complicado, entonces debemos preguntar cómo se formó un algoritmo de eficacia tan extraordinaria. La respuesta normal, por supuesto, sería que surgió por “selección natural”. Por lo que respecta a las criaturas con cerebros evolucionados, aquellos con los algoritmos más eficaces tendrían más probabilidades de sobrevivir y por ello, en general, tendrían más prole. Esta prole también tendería a portar algoritmos más eficaces que sus padres, puesto que heredaron de sus padres los ingredientes de estos mejores algoritmos; de este modo los algoritmos mejoraron poco a poco —no necesariamente de manera uniforme puesto que pudo haber tropiezos importantes en su evolución— hasta que alcanzaron el importante estado que (aparentemente) encontramos en el cerebro humano. Incluso según mi propio punto de vista tendría que haber algo de verdad en esa imagen, puesto que concibo que gran parte de la acción del cerebro es realmente algorítmica y, como el lector habrá deducido de lo anterior, creo firmemente en el poder de la selección natural. Pero no veo cómo la selección natural por sí sola pueda hacer evolucionar algoritmos que pudieran tener el tipo de juicios conscientes sobre la validez de otros algoritmos que al parecer tenemos. Imaginemos un programa ordinario de computadora. ¿Cómo llegó a formarse? Es evidente que no (directamente) por selección natural. Algún programador humano de computadoras lo habrá concebido, verificando que realiza correctamente las acciones que se supone debe hacer. (En realidad, muchos programas de computadora complicados

En este estado de cosas, podría decirse que, aunque la IA pueda hacer muchas y grandes cosas, no puede pensar, y mucho menos ser consciente de sí misma, porque, aun cuando parezca que está pensando, en realidad, está empleando una técnica conocida como comparación de patrones (Copeland,

---

contienen errores —normalmente menores, pero a menudo muy sutiles y que no salen a la luz excepto en circunstancias muy poco comunes—. La presencia de tales errores no afecta medularmente a mi argumento.) A veces un programa de computadora puede haber sido “escrito” por otro programa, digamos un programa de computadora “maestro”, pero en tal caso el propio programa maestro habrá sido el producto del ingenio y la intuición humanos; o el programa podría perfectamente ensamblarse a partir de ingredientes, algunos de los cuales son los productos de otros programas de computadora. Pero en todos los casos la validez y la misma concepción del programa habrá sido en última instancia responsabilidad de (al menos) una conciencia humana. Podemos imaginar, por supuesto, que no es necesario que haya sido así y que, dado el tiempo suficiente, el programa de computadora pudo haber evolucionado espontáneamente por algún proceso de selección natural. Si creemos que las acciones de las conciencias de los programadores de computadoras son en sí mismas simples algoritmos, entonces debemos creer que los algoritmos han evolucionado de esta misma forma. Lo que me molesta de esto, sin embargo, es que la decisión sobre la validez de un algoritmo no es en sí misma un proceso algorítmico [...]. Para decidir si un algoritmo funcionará o no, necesitamos perspicacia, y no sólo otro algoritmo. De todas formas, aun sería posible imaginar algún tipo de proceso de selección natural que fuera efectivo para producir algoritmos aproximadamente válidos. Sin embargo, yo personalmente encuentro esto muy difícil de creer. El escaso conocimiento que he adquirido acerca de cómo funcionan los cerebros humanos —y, de hecho, cualquier otra cosa viviente— me deja perplejo de asombro y admiración. El funcionamiento de una neurona es extraordinario, pero las propias neuronas están organizadas en conjunto, en el instante del nacimiento, de una forma muy notable con un gran número de conexiones dispuestas para todas las tareas que sean necesarias más adelante. No es sólo la propia conciencia lo que es notable, sino todo lo que parece ser necesario para sostenerla. El mensaje debería estar claro. La verdad matemática no es algo que adivinemos simplemente utilizando un algoritmo. Creo, también, que nuestra conciencia es un ingrediente fundamental en nuestra comprensión de la verdad matemática. Debemos “ver” la verdad de un argumento matemático para estar convencidos de su validez. Esta “visión” es la esencia misma de la conciencia. Como he declarado antes, una buena parte de las razones para creer que la conciencia es capaz de influir en los juicios de verdad de una manera no algorítmica se derivan del teorema de Gödel. Si podemos ver que el papel de la conciencia es no algorítmico cuando formamos juicios matemáticos, en lo que el cálculo y la demostración rigurosa constituyen un factor tan importante, entonces podremos persuadirnos de que un ingrediente no algorítmico semejante podría ser también crucial para el papel de la conciencia en circunstancias más generales (no matemáticas)».

1993: 75). Aunque una máquina pueda realizar acciones que podrían ser consideradas propias de un ser humano, ello responde a una mera gestión de símbolos carente de toda comprensión, a diferencia de lo que ocurre en el caso de los seres humanos (*ibid.*: 40). Tal y como han puesto de manifiesto expertos en la materia, los robots, del tipo que sean, son meros autómatas (Ortego Ruiz, 2022: 67). El experimento de la «habitación china» permite llegar a esa conclusión<sup>18</sup>. No hay forma de escapar de la habitación desarrollando programas de *software* basados en la mera computación. El robot lo único que hace es manipular signos, y por ese camino no puede saltar de ningún modo a la genuina comprensión humana. Para crear una inteligencia artificial fuerte no basta con aumentar la complejidad de los programas y de los procesadores, ya que añadir más de lo mismo no soluciona el problema (Navarro Reyes, 2005: 267).

Finalmente, tal y como afirman destacados neurocientíficos, atendiendo a la cuestión central de este epígrafe, es importante señalar que no hay razón para pensar que la IA, solo porque se esté volviendo más inteligente, se volverá consciente<sup>19</sup>. Esta reciente afirmación puede hacer que las palabras pronunciadas por el profesor Jefferson en 1949, en la disertación de la medalla de Lister, dejen de

<sup>18</sup> «Dentro de la habitación existe un mecanismo que desconoce por completo el idioma chino y que se encuentra aislado del exterior. Solo cuenta con reglas dadas y cerradas que le indican qué sinograma emitir en función del que entra. En virtud de los resultados podría decirse que el sistema dentro de la habitación china conoce el idioma chino, y de ser un robot podría parecer que es inteligente. Sin embargo, la conclusión es diferente. El mecanismo o robot dentro de la habitación conoce la síntesis del chino, pero no la semántica de dicho idioma, su contenido. Por lo tanto, no puede concluirse que entienda chino» (véanse Searle, 2001: 37; Navarro Reyes, 2005: 268; Ortega Ruiz, 2022: 67).

<sup>19</sup> Véase Seth (2023: 232), profesor británico de Neurociencia Cognitiva de la Universidad de Sussex, quien considera que nadie sabe con certeza cómo es que nuestros cerebros generan sueños y deseos, placer y dolor, y por qué, aunque todos tengamos una biología similar, creamos experiencias subjetivas muy distintas a las de otra persona. Para este problema difícil, ninguna explicación podría estar jamás a la altura de resolverlo. En esta misma línea, Chopra (2020: 50), reconocido escritor y conferenciante, licenciado en Medicina por el All India Institute of Medical Science, especializado en endocrinología, profesor de Medicina en la Universidad de Boston, ha debatido y conversado con muchos neurocientíficos y «ninguno ha sido capaz de responder a las preguntas más simples sobre la conciencia, entre las que se incluyen: ¿qué es un pensamiento? ¿De qué forma la actividad electroquímica de una neurona se convierte en palabras, visiones y sonidos en nuestras cabezas? ¿Por qué el siguiente pensamiento de una persona es impredecible? Si alguien tiene un vocabulario de 30 000 palabras, ¿significa esto que un grupo de células cerebrales conoce 30 000 palabras? En caso afirmativo, ¿de qué manera se almacena la palabra? Para la palabra *gato* ¿hay un lugar dentro de una célula cerebral que contenga las letras g-a-t-o? Nadie puede responder adecuadamente a ninguna de estas preguntas».

considerarse mera expresión de solipsismo (Turing, 1950: 12) y se revelen como lo que verdaderamente son, una manera de evidenciar que la diferencia entre los seres humanos y los robots se asienta en algo que está más allá de lo que es meramente tangible o medible, *la conciencia y la consciencia de la experiencia en su totalidad*:

Hasta que una máquina pueda escribir un soneto o componer un concierto debido a las emociones y pensamientos que tuvo, y que no sea debido al uso de símbolos al azar, podremos estar de acuerdo que máquina es igual a cerebro, es decir, no solo que lo escriban, sino saber que lo escribió. Ningún mecanismo podría sentir placer por sus éxitos, sentir pesar cuando se le funde una válvula, sentirse bien con un halago, sentirse miserable por sus errores, estar encantado por el sexo, estar enojado o deprimido cuando no consigue lo que quiere (*ibid.*: 11).

Como dice Timnit Gebru: «¿Queremos algo que escriba nuevos libros al azar basados en combinaciones del conjunto de libros que ya existen?». Ella no lo cree, y es probable que la gran mayoría tampoco<sup>20</sup>.

## 2. HACIA UN AVANCE RESPONSABLE DE LA INTELIGENCIA ARTIFICIAL

Conocer la naturaleza exclusivamente algorítmica de la IA, aunque aún no pueda determinarse con precisión hasta dónde podrá llegar, es el punto de partida necesario para poder determinar, con el mayor acierto posible, cuál deba ser el régimen jurídico que con solvencia responda a toda la problemática que se suscita en torno a las máquinas y su integración en la vida diaria. Máxime si con ello apoyamos todos los esfuerzos que los expertos en IA están realizando para frenar la carrera sin control de estos algoritmos con aspecto humano. Es importante no perder de vista que los llamados «humanoides» son meramente «cosas»<sup>21</sup>, y, como tales, son sencillamente objeto del derecho, y no sujetos de derechos<sup>22</sup>.

<sup>20</sup> Véase Sánchez Díez (2023a: 16). Timnit Gebru dirigía el equipo de ética de la inteligencia artificial en Google. En 2020, tras publicar un artículo donde advertía de los peligros del desarrollo sin control, ni regulación de la inteligencia artificial, fue despedida. Científica etíope, experta en computación, especializada en algoritmos de minería de datos y sesgo algorítmico.

<sup>21</sup> Y no me refiero a «otra cosa», como se ha puesto de manifiesto por Gavilán (2023: 521), considerando que deben tener un tratamiento jurídicamente diferenciado con respecto a la inteligencia artificial en general. Es cierto que presentan características diferenciales como la corporeidad, el movimiento, la aparente personalidad, la posibilidad de interacción de tú a tú, la individualización..., pero no nos confundamos, son meras «cosas», y, como tales, su régimen jurídico se ha de adecuar a ello.

<sup>22</sup> Los robots no dejan de ser máquinas o herramientas, como queramos llamarlo, es decir, cosas al servicio del hombre, que pretenden mejorar y facilitar su vida. Como

Estas herramientas de aprendizaje autónomo, que emulan la inteligencia humana, están absorbiendo millones de textos de un internet donde imperan puntos de vista racistas y misóginos. Lo que hacen es derramar una visión del mundo sesgada que puede perpetuar y amplificar la injusticia y la discriminación. Los derechos fundamentales y su protección se están viendo seriamente afectados, y es importante establecer los límites legales necesarios para garantizar su libre y adecuado ejercicio<sup>23</sup>. Los efectos discriminatorios que atentan contra el derecho fundamental a la igualdad, reconocido en el art. 14 de la Constitución, se están dejando sentir en el empleo y en el ámbito sanitario, a la vez que, de manera paralela y como consecuencia de ello, también se está atentando contra el derecho fundamental a la integridad física<sup>24</sup>. Amazon, una

---

dice López de Mántaras (López de Mántaras y Fernández Burgueño, 2019: 1): «[...] hay que mantener a las máquinas separadas del ser humano, no tenemos que verlas como sustitutos, sino como colaboradores, *coworkers* y herramientas... La IA será un colaborador o partner del ser humano. Podrá automatizar ciertas tareas, pero no desplazar a la persona de un puesto de trabajo». En esta misma línea, Fernández Burgueño, jurista y docente especializado en legaltech, metaverso y ciberseguridad, considera que «la entrada de este tipo de herramientas va a ser progresiva, por lo que nos va a ayudar a ir adaptándonos para poder hacer las cosas mejor, más rápido y de manera más eficiente» (*ibid.*: 2).

<sup>23</sup> Véase Sánchez Díez (2023b: 58): «En febrero de 2019, Nijeer Parks, un carpintero de la ciudad de Patterson, en Nueva Jersey, recibió una llamada de su abuela: la policía de la localidad se había presentado en la casa que ambos compartían para arrestarlo, acusado de un robo que había terminado con su huida y el intento de atropello de los agentes de policía. Park pensó que se trataba de un error sin importancia [...], pero pasó los siguientes diez días en prisión». Una vez liberado entendió lo que había pasado, «una herramienta de reconocimiento facial lo había señalado como responsable», lo que, además, pone de manifiesto que «la detención del propio Nijeer Parks es un ejemplo de cómo datos corrompidos pueden llevar a errores de consecuencias catastróficas. El suyo es uno de los cinco casos conocidos de detenciones policiales erróneas basadas en sistemas de reconocimiento facial en Estados Unidos, y los cinco hombres son negros. Ya desde que salió en 2015 a la luz que el algoritmo de Google Photos etiquetaba como Gorilas los rostros de personas negras, los sistemas de reconocimiento facial alimentados por inteligencia artificial han demostrado una y otra vez que son más propensos a cometer equivocaciones cuando se trata de reconocer e identificar rostros de color».

<sup>24</sup> Véase Sánchez Díez (2023b: 60): «Los algoritmos también pueden dictar qué tipos de cuidados médicos recibimos. En 2019, un grupo de investigadores de la Universidad de Berkeley (California) descubrió que existía un sesgo racial en uno de los algoritmos más extendidos en los hospitales estadounidenses, un modelo matemático que se aplica a más de 200 millones de personas al año. El algoritmo, en su

de las corporaciones líderes en el uso de la inteligencia artificial, descubrió en 2015 que el sistema automático que se estaba empleando por la compañía para examinar candidatos discriminaba a las mujeres. Los propios datos justificaban el error. El sistema estaba aplicando un patrón basado en los datos recopilados y en el volumen masivo de información que no filtra los sesgos, ni los detecta, y mucho menos tiene capacidad para hacerlo<sup>25</sup>.

Es un hecho constatado que la IA se ha extendido muy rápido por todas partes. Así lo expresan numerosos expertos en la materia. Hace apenas un par de años ya se hablaba de los peligros de los modelos del lenguaje y de los sesgos, aunque ese momento nadie sabía lo que eran, y ahora cada ser humano está hablando del ChatGPT. Aparece en las noticias de todo el mundo. Los motores de búsqueda no dejan claro que hay diferentes puntos de vista, o que detrás de esa información que se está obteniendo hay personas con su manera de ver y entender el mundo. Tampoco te da la opción de que examines las diferentes fuentes para que decidas con criterio propio. El sistema te hace creer que tiene respuesta para todo, cuando, en realidad, la respuesta que está dando puede ser totalmente errónea, y el usuario va a tomar decisiones basadas en ella<sup>26</sup>.

Esta velocidad en la proliferación de la inteligencia artificial ha llevado a numerosos expertos del panorama internacional a firmar una misiva pidiendo el freno temporal en su desarrollo para evaluar los riesgos. De la misma manera, otro sector, no menos representativo, lo considera innecesario y alarmista<sup>27</sup>.

---

intento de asignar recursos sanitarios de forma eficiente, ponderaba el coste por individuo como variable para clasificar a los pacientes en función de su gravedad. Pero como el sistema sanitario tradicional gasta menos en afroamericanos que en los blancos, consideraba erróneamente que los pacientes negros están más sanos y no requerían atención adicional, aunque en realidad padecieran los mismos o peores problemas de salud».

<sup>25</sup> Véase Sánchez Díez (2023b: 60): «Históricamente Amazon ha contratado a menos mujeres que hombres, con independencia de sus cualificaciones [...], asignaba una puntuación menor a currículums que mencionaran la palabra “mujer”, como por ejemplo haber sido la capitana de un club de ajedrez femenino».

<sup>26</sup> Véase Sánchez Díez (2023a: 18): «Queremos ir hacia sistemas que eduquen al público en que hay distintas formas de obtener información, en lugar de fingir que tienes respuesta para todos». Esta es la respuesta de Timnit Gebru cuando la periodista le pregunta qué es lo que más le preocupa sobre la aplicación de la inteligencia artificial a los motores de búsqueda como Google, de los que tantos de nosotros dependemos para encontrar información básica.

<sup>27</sup> Según expone para *Computer* la periodista González Valenzuela (2023): «Algunos líderes como Sundar Pichai, CEO de Google, o Sam Altman, CEO de OpenAI, han dicho que no cumplirán con el llamado a hacer una pausa, afirmando que apoyan la



Ahora bien, lo que sí que parece importante y urgente es determinar con claridad y precisión el régimen jurídico aplicable a las situaciones que ya son una realidad y que causan daño a los usuarios o atentan contra sus derechos fundamentales. Atribuir responsabilidades en un marco jurídico adecuado, probablemente, ralentizará el avance de la inteligencia artificial de manera espontánea y responsable, sin necesidad de esfuerzos añadidos. Quizás esta sea, en gran medida, la razón por la que existen voces que se alzan en contra de la regulación, o pretenden controlarla, considerando que puede suponer un freno a la innovación, cuando en realidad lo que pretenden eludir es su responsabilidad<sup>28</sup>.

## II. LA PERSONALIDAD JURÍDICA DE LOS ROBOTS: ¿UNA REALIDAD JURÍDICA PLAUSIBLE Y NECESARIA?

La primera vez que escuché hablar de la posibilidad de asignar personalidad jurídica a los robots he de reconocer que me escandalicé. Me parecía una idea absurda y carente de sentido. Solo se me ocurrían argumentos en contra de semejante posibilidad. En un primer momento pensé que, como

---

regulación en lugar de una pausa. Por otro lado, Andrew Ng, fundador y CEO de Landing AI y Google Brain —considerado por muchos como el gran gurú y experto del sector— se posiciona claramente en este aspecto muy en contra del parón: *La IA responsable es importante y la IA tiene riesgos. La narrativa popular de la prensa de que las empresas de IA se están volviendo locas enviando códigos inseguros no es cierta. La gran mayoría (lamentablemente, no todos) de los equipos de IA se toman en serio la IA responsable y la seguridad. Invirtamos más en seguridad mientras avanzamos en la tecnología, en lugar de sofocar el progreso [...]. Una moratoria de 6 meses no es una propuesta práctica. Para avanzar en la seguridad de la IA, las regulaciones sobre transparencia y auditoría serían más prácticas y marcarían una mayor diferencia».*

<sup>28</sup> Según expone Perrigo (2023), aunque realmente Sam Altman, CEO de OpenAI, habló repetidamente de la necesidad de una regulación global de la IA, por detrás de escena, OpenAI, trata de controlar elementos importantes de la legislación de la IA más completa del mundo (la ley de la IA de la UE), para que se diluyan de manera que reduzcan la carga regulatoria para la empresa. En varios casos OpenAI ha propuesto enmiendas que luego se hicieron al texto final de la ley de la UE, que fue aprobada por el Parlamento Europeo el 14 de junio de 2023. En 2022, OpenAI argumentó repetidamente ante los funcionarios europeos que la próxima ley de IA no debería considerar sus sistemas de IA de propósito general, incluidos GPT-3, el precursor de ChatGPT, y el generador de imágenes Dall-E 2, como de «alto riesgo», una designación que los sometería a estrictos requisitos legales que incluyen transparencia, trazabilidad y supervisión humana.

categoría jurídica, era una empresa más que cuestionable. Pero, a medida que iba leyendo los distintos argumentos esgrimidos por autorizada doctrina, comencé a abrirme a la posibilidad de su adopción, e incluso a considerar cada planteamiento de lo más razonable. Actualmente, no me cabe duda de que la personalidad jurídica de los robots es una realidad plausible. Sin embargo, hay una pregunta que late con fuerza en mi interior, una pregunta en la que descansa mi humilde aportación: «¿Para qué?». ¿Necesitamos realmente atribuir personalidad jurídica a los robots para dar respuesta a los interrogantes que se suscitan en torno a su regulación, o no es más que una excusa para eludir responsabilidades al abrigo de una «tergiversada innovación»?

## 1. LA PERSONALIDAD JURÍDICA DE LOS ROBOTS COMO CATEGORÍA JURÍDICA<sup>29</sup>

Con independencia de la existencia de posturas enfrentadas acerca de posibilidad de dotar o no de personalidad jurídica a los robots, ha quedado suficientemente constatado y argumentado que esta categoría jurídica, como tal, es perfectamente posible<sup>30</sup>. Es decir, es algo que el derecho puede hacer, igual que con anterioridad ha ocurrido en otros casos precedentes, como el de las personas jurídicas<sup>31</sup>. En este sentido, puede decirse que históricamente el derecho ha relativizado el vínculo ser humano-derecho, unas veces por defecto, negando tal condición a los esclavos, y otras por exceso, atribuyendo tal condición a grupos de personas, masas de bienes u organizaciones (Sánchez-Castro Díaz-Guerra, 2023: 1).

Es cierto que la carta pública dirigida a la Comisión Europea de Inteligencia Artificial y Robótica<sup>32</sup>, firmada por más de doscientos expertos europeos, procedentes de una diversidad de ámbitos que van desde el tecnológico y legal

<sup>29</sup> Véase Núñez Zorrilla (2019: 55). Es importante resaltar que hay quien considera que, por muy compleja y avanzada que sea la inteligencia de los robots, no dejan de ser lo que son, «máquinas», «cosas» que pueden ser desconectadas en cualquier momento; otra cosa distinta es que, sin dejar de ser cosas, se les atribuya un estatus jurídico específico o especial dentro de esta categoría, con la finalidad de proteger determinados intereses de la sociedad o de las auténticas personas. Véase Rogel Vide (2018: 87): en cualquier caso, sean o no considerados de acuerdo con verdadera naturaleza, podría decirse que no está justificado asignarles personalidad jurídica.

<sup>30</sup> Véase Macanás (2023: 31): «No hay norma, principio o sistema que haga imposible la aplicación de la categoría jurídica de personalidad a entes no humanos».

<sup>31</sup> Véanse Sánchez del Campo (2016: 10); Díaz Calvarro (2021: 57); Sánchez-Castro Díaz-Guerra (2023: 7); De Castro y Bravo (1981: 137).

<sup>32</sup> *Open Letter to the European Commission Artificial Intelligence and Robotics*.

hasta el sanitario y ético, ha pedido que no se reconozca personalidad jurídica a los robots<sup>33</sup>. Pero en ningún caso duda sobre la posibilidad legal de adoptar esta categoría jurídica (Fernández y Cortés, 2018: 2).

Son numerosas las voces que se han alzado en contra de esta tendencia de dotar de personalidad jurídica a los robots<sup>34</sup>, al mismo tiempo que, en ocasiones, ha podido parecer mayoritaria la postura contraria<sup>35</sup>. Incluso se han podido apreciar posiciones que no se mostraban fanáticas de ninguna de las dos opciones (Sánchez del Campo Redonet, 2016: 54). Por su parte, el Parlamento Europeo ha evolucionado en esta línea, desde la posibilidad de reconsiderarlo como una opción y lanzarlo como propuesta a los Estados miembros<sup>36</sup> hasta llegar a la situación actual, que trae causa de la Resolución

<sup>33</sup> Véase Fernández y Cortés (2018: 1): «Los firmantes del documento se muestran en principio favorables a que la Unión Europea establezca una regulación de la robótica y la inteligencia artificial adecuada para garantizar un alto nivel de certeza y seguridad de los ciudadanos europeos que, a la vez, favorezca la innovación. Sin embargo, esos mismos firmantes del documento muestran su preocupación por la recomendación contenida del documento del Parlamento Europeo que propone que se atribuya una personalidad electrónica en aquellos casos en los que los robots adopten decisiones autónomas o interactúen independientemente con terceros».

<sup>34</sup> Entre otros, Natali Navejans, profesora de Derecho en la Universidad de Artois de Francia y principal impulsora de la carta, Noel Sharket, profesor emérito de Inteligencia Artificial y Robótica en la Universidad Sheffield, Ramón López de Mántaras, profesor de Investigación del CSIC y director del Instituto de Investigación en IA, y Jean-Claude Juncker, cuando fue presidente de la Comisión Europea (véanse Díaz Calvarro, 2021: 57; Fernández y Cortés, 2018: 3).

<sup>35</sup> A favor de una personalidad jurídica, encontramos destacadas figuras del panorama internacional y nacional, entre otros, Mady Delvaux-Stheres, cuando era diputada al Parlamento Europeo, redactora de la controvertida resolución europea para dotar de personalidad jurídica a los robots, Pablo García Mexía, consultor-director de derecho digital en Herbert Smith Freehills, Madrid, letrado de las Cortes Generales, codirector del posgrado en Privacidad y Sociedad Digital de la Universidad Autónoma de Madrid, director de la revista de *Privacidad y Derecho Digital*, y Moisés Barrios Andrés, letrado del Consejo de Estado, profesor de Derecho Digital, árbitro y abogado especializado en nuevas tecnologías (véanse Ortego Ruiz, 2022: 73; Fernández y Cortés, 2018: 2; Aransay Alejandre, 2019:107).

<sup>36</sup> «En el Proyecto de informe con recomendaciones a la Comisión sobre normas de Derecho Civil relativas a la robótica, publicado el 31 de mayo de 2016, se afirmaba que procedía comenzar por las cuestiones relativas a la responsabilidad civil, y para solucionar esta cuestión desde un punto de vista técnico-jurídico, se reputaba cada vez más urgente abordar la cuestión fundamental de si los robots deben tener personalidad jurídica. Lo que inevitablemente abría otra cuestión, cual es el la 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 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 daños. No hace falta insistir en el calado de esas aseveraciones pese a que, como veremos, no hizo suyas finalmente el PE en su integridad, aunque sí asumió la necesidad de estudiar la cuestión y resolverla a futuro. Todo ello, en contra de un Informe de la Comisión de Asuntos Jurídicos del Parlamento Europeo, sobre la evaluación y análisis, bajo las perspectivas jurídica y ética, de las futuras reglas europeas de derecho civil sobre robótica» (véase Coca Payeras, 2023: 10-11). «La Resolución del PE de 16 de febrero de 2017 con recomendaciones a la Comisión sobre Normas de Derecho Civil sobre robótica. En la Introducción se reproducían algunos de los considerandos ya reseñados del Informe de 27 enero 2017 y en el apartado de Responsabilidad, epígrafe 59, se sigue pidiendo a la Comisión que, cuando realice una evaluación de impacto de su futuro instrumento legislativo, explore, analice y considere las implicaciones de todas las posibles soluciones jurídicas, y entre ellas en el apartado f se reitera el texto del Informe: crear a largo plazo una personalidad jurídica específica para los robots, 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» (véanse Fernández y Cortés, 2018: 2; Coca Payeras, 2023: 12). Por otro lado, «el Informe de la Comisión de Asuntos Jurídicos del Parlamento Europeo, sobre la evaluación y análisis, bajo las perspectivas jurídica y ética, de las futuras reglas europeas de derecho civil sobre robótica, suscrito por la Profesora Nathalie Nevejans, en su apartado 3.1 puso de relieve la inutilidad e incongruencia de pretender solucionar la cuestión de la responsabilidad civil de los robots dotándoles de personalidad, creando las personas electrónicas. Se indicaba que “desde un punto de vista científico, jurídico e incluso ético, es hoy imposible, y probablemente lo sea aún por mucho tiempo, que un robot pueda participar en la vida jurídica sin la presencia de un ser humano oculto tras él”. Y se enfatizaba cómo “ese estatuto de persona generaría necesariamente efectos jurídicos indeseables como el reconocimiento de derechos y deberes a una simple máquina. ¿Cómo podría un robot tener deberes, siendo que tal noción está estrechamente ligada a las representaciones morales humanas? Y aún más, ¿qué derechos se atribuirían a los robots?: El derecho a la vida —o sea, el derecho a la no destrucción—; el derecho a la dignidad, a la igualdad con el ser humano, a una retribución por su trabajo...”. Y seguía: “En realidad, los partidarios de la personalidad jurídica de los robots tienen una visión fantasiosa del robot, influenciada por las novelas o el cine de ciencia ficción. No pueden concebir el robot —sobre todo si es calificado de inteligente o humanoide— más que como una criatura artificial pensante, un *alter ego* del ser humano”. Por otro lado, advertía: “El peligro no estriba únicamente en reconocer derecho y obligaciones a una simple útil, sino también en hacer estallar las fronteras entre el ser humano y la máquina, abriendo el camino a la confusión entre lo vivo y lo inerte, entre lo humano y lo inhumano [...] poniendo en

del Parlamento Europeo de 20 de octubre de 2020, con recomendaciones a la Comisión, para un régimen de responsabilidad civil en materia de inteligencia artificial: «Todas las actividades, dispositivos o procesos físicos o virtuales gobernados por sistemas de IA pueden ser técnicamente la causa directa o indirecta de un daño o un perjuicio, pero casi siempre son el resultado de que alguien ha construido o desplegado los sistemas o interferido en ellos; observa, este respecto, que no es necesario atribuir personalidad jurídica a los sistemas de IA»<sup>37</sup>.

Por un lado, este es el punto de partida de la propuesta de reglamento denominada «Ley de Inteligencia Artificial», que lleva a cabo una regulación general de gestión de riesgos para normas sobre IA armonizadas y, por otro, diversas iniciativas, que también se hallan en curso, relativas a la regulación de la responsabilidad civil por los daños causados por sistemas de IA, como son la propuesta de directiva denominada «Directiva de responsabilidad por IA», que tiene por objeto adaptar determinados aspectos de la prueba de la responsabilidad extracontractual por culpa a los requerimientos de la inteligencia artificial, y la propuesta de directiva de responsabilidad por productos defectuosos, que deberá sustituir a la Directiva 85/374/CEE sobre la misma materia, para facilitar su aplicación a los sistemas de IA (Martín Casals, 2023: 59-60; Arrellano Toledo, 2023; Rebollo Delgado, 2023: 37), que, finalmente no contemplan la posibilidad de dotar a los sistemas de aprendizaje autónomo de la IA de ningún estatuto jurídico que tenga que ver con la personalidad jurídica o categoría intermedia.

Las distintas propuestas, sobre las que se va a incidir en el último epígrafe de este trabajo, garantizan que la IA se utilice de manera ética y segura, protegiendo los derechos y valores fundamentales de los ciudadanos europeos. Para ello siguen un enfoque de regulación de riesgos que aborda riesgos como los posibles sesgos, errores y opacidad que pueden afectar negativamente a los derechos fundamentales de los ciudadanos, y establece una serie de requisitos obligatorios para el diseño y desarrollo de sistemas de IA antes de su comercialización, y su posterior seguimiento y supervisión (Martín Casals, 2023: 60).

---

cuestión los fundamentos humanistas de Europa”. Y concluía en este punto: “Otorgar el estatus de persona a un ente no viviente y no consciente sería un error y tal solución conllevaría el riesgo a largo plazo de reducir al ser humano al rango de simple máquina. Los robots deben ocupar un lugar al servicio de la humanidad y no deberían ocupar cualquier otro, salvo en la ciencia ficción” (véanse Fernández y Cortés, 2018: 2; Coca Payeras, 2023: 14-15).

<sup>37</sup> Resolución del Parlamento Europeo de 20 de octubre de 2020, con recomendaciones a la Comisión, para un régimen de responsabilidad civil en materia de inteligencia artificial. Introducción, epígrafe 7.

## 2. LA PERSONALIDAD JURÍDICA DE LOS ROBOTS Y LA «FALACIA» DE SU NECESIDAD

Aunque no sea imposible adoptar para los robots el estatuto de la personalidad jurídica en nuestro ordenamiento (y en otros), aunque se considere que lo realmente importante para dotar a un ente no humano de personalidad jurídica sea la existencia de un interés jurídico relevante (Macanás, 2023: 39), no se debe ignorar la verdadera «necesidad», o no, de semejante empresa. Ciertamente, este *statu quo* es una realidad jurídica plausible, e, igualmente, responde a un interés jurídico concreto, que va más allá del entusiasmo de la innovación o de la solución novedosa (*ibid.*: 42). Sin embargo, aunque se den ambos requisitos, carecerían de sentido todo el esfuerzo normativo y su justificación, si finalmente no fuera necesario crearlo *ex novo*, por tenerlo sencillamente en nuestras manos. En este sentido, hay quien considera un apriorismo falaz la asunción de que las nuevas tecnologías basadas en la IA, por el simple hecho de ser nuevas, no tienen reflejo exacto en el régimen jurídico de la responsabilidad civil<sup>38</sup>.

En esta línea, resulta de especial interés asomarse a la propuesta de directiva europea para establecer la responsabilidad por daños causados por sistemas de inteligencia artificial, que se podría empezar a aplicar durante el año 2024. El titular de la noticia que se hace eco del XXV Congreso de Responsabilidad Civil y Seguro no deja lugar a duda sobre la filosofía de esta propuesta: «No habrá régimen especial para la responsabilidad por daños causados por la IA»<sup>39</sup>.

---

<sup>38</sup> Véase Ataz López (2020: 13), quien, además, considera que «una de las grandes virtudes de nuestro código es la generalidad de sus formulaciones que permite a muchos preceptos reinventarse a lo largo del tiempo sin necesidad de su modificación. En particular eso ocurre con el artículo 1902 del CC, que, nacido para una sociedad rural y poco tecnificada, en la que los daños que una persona podía causar a otra eran muy limitados, como demuestra la relativa poca importancia que jurisprudencia y doctrinalmente tenía la cuestión de la responsabilidad civil a finales del siglo XIX, sigue hoy prestando sus servicios para la generalidad de los supuestos. Es cierto que ha habido que modificar los criterios jurisprudenciales, relativos a la información del estándar de la diligencia exigible en algunos casos, y sobre todo, revisar las cuestiones relacionadas con la carga de la prueba, pero al final el artículo 1902 permanece inmodificado, y sigue siendo tan útil en nuestra post-industrial como lo fue en la sociedad rural española del último cuarto del siglo XIX, o incluso más útil hoy día, puesto que el número de supuestos en los que resulta de aplicación ha aumentado».

<sup>39</sup> «No habrá un régimen especial para la IA, lo que hay son aportaciones a regímenes especiales, como, por ejemplo, el de productos defectuosos» (véase Rodríguez de las Heras, 2023: 2).

Igualmente, no se puede aspirar a crear un nuevo derecho fundamental como consecuencia del surgimiento de cada una de las nuevas posibilidades tecnológicas, ya que no todas tienen el mismo grado de incidencia ni la misma trascendencia desde una perspectiva jurídica, incluso los denominados «derechos digitales» por la LO 3/2018 son una concreción o variante aplicativa de los derechos fundamentales ya reconocidos en nuestro ordenamiento jurídico y en el europeo, y suficientemente garantizados<sup>40</sup>. Hasta ahora, ha sido a través de la legislación y la jurisprudencia como se han ido adecuando los derechos existentes a las nuevas vulneraciones, ampliando las interpretaciones y solventando las omisiones, a la vez que formulando tendencias interpretativas, que con el tiempo tienen su plasmación en nuevas normativas<sup>41</sup>.

En este sentido, tal y como se ha puesto de manifiesto, la primera declaración que ha de hacerse, ante la emergencia de la sociedad digital, ya sea a

<sup>40</sup> Véase Rebollo Delgado (2023: 51). Igualmente, es importante recordar que la Carta de los derechos digitales, de 14 de julio de 2021, no tenía como propósito crear nuevos derechos fundamentales: «El propósito de la Carta, ya se sabe, no es crear nuevos derechos fundamentales sino perfilar los ya reconocidos constitucionalmente y adecuar su protección al entorno digital. En cuanto a su eficacia jurídica, pese a carecer de naturaleza normativa sino meramente programática, se erige en marco de referencia para los poderes públicos en relación con los retos en el entorno digital, sirviendo de referencia para futuros proyectos legislativos y políticas públicas en España enderezadas a la consecución de ese fin» (González de Patto, 2023: 385).

<sup>41</sup> Véase Rebollo Delgado (2023: 5), que, además, en este sentido, considera muy clarificadora, a este respecto, la STC 58/2018: «Los avances tecnológicos y el fenómeno de la globalización a través de internet y de otras vías dan lugar a nuevas realidades que, de una u otra forma, pueden incidir sobre el ejercicio de los derechos fundamentales, su delimitación, y su protección, lo que obliga a este tribunal a una constante labor de actualización de su doctrina para adecuarla a la cambiante realidad social, con el fin de dar una repuesta constitucional a esas nuevas situaciones». De igual forma, continúa exponiendo este autor: «[...] la jurisprudencia habrá de ir tejiendo la argumentación y fundamentación de variaciones en los derechos, armando una estructura lógico-jurídica de los derechos fundamentales, que produzca una respuesta adecuada a las nuevas necesidades jurídicas y que palie las deficiencias normativas, a la vez que facilite su aplicabilidad y su ejecución material. La protección jurídica y los medios de garantía ya existen, lo que varían son las formas de vulneración, por ello no es necesario crear un ordenamiento jurídico *ex novo*, es necesario ir adecuando el existente a las nuevas necesidades, y ello es tarea tanto del legislador, de la jurisprudencia y de la doctrina». En cualquier caso, hay quien considera que se deberían reconocer nuevos derechos fundamentales *stricture sensu* en las constituciones, como el acceso universal a internet y la ciberseguridad, así como un derecho fundamental a la verdad para combatir las noticias falsas (*fake news*), y un derecho a la conciliación familiar y laboral (véase Barrios Andrés, 2021: 17).

nivel constitucional o legal, debe consistir en un compromiso doble. Por una parte, el de potenciar al máximo, siempre en beneficio de los ciudadanos, el desarrollo, uso y empleo de la sociedad digital con todos sus medios e instrumentos. Por otra, el de garantizar sin desfallecimiento todos los principios y valores fundamentales de nuestra convivencia recogidos en el art. primero de la Constitución, y a lo largo de toda ella, así como la dignidad de la persona humana con los derechos fundamentales que le son inherentes<sup>42</sup>.

---

<sup>42</sup> Véase De la Quadra Salcedo Fernández del Castillo (2018: 66), quien, además, expresa que «la igualdad es uno de esos derechos fundamentales que pueden quedar comprometidos con la aparición de técnicas que permitan el llamado aumento o potenciación de determinadas personas mediante su conexión a ordenadores con inteligencia artificial, lo que puede determinar la competencia en el mercado, incluido el mercado de trabajo, en condiciones de igualdad». De hecho, como ya se expuso en apartados anteriores, los efectos discriminatorios de la inteligencia artificial se han dejado sentir en el empleo y en otros ámbitos. En esta misma línea, García-Antón Palacios (2023: 173-174) expone que «uno de los retos actuales a los que se enfrenta la IA es la posibilidad de que los algoritmos de los sistemas de aprendizaje automático sean entrenados con valores que reflejen ciertos sesgos discriminatorios como el sexo, la raza, la edad, la religión, o cualquier otro. Para hacer frente a las posibles situaciones de discriminación y vulneración de los derechos fundamentales, la “Rome Call for AI Ethics” indica que la libertad y la dignidad humanas descritas en los artículos 1 y 2 de la DUDH deben protegerse y garantizarse al producir y utilizar estos sistemas, por lo que los mismos deben concebirse, diseñarse e implementarse para servir y proteger a los seres humanos y el entorno en el que viven [...]. A juicio de la Agencia Europea para los Derechos Fundamentales, en consonancia con el derecho a la igualdad y la prohibición de discriminación de los artículos 20 y 21 de la Carta, la discriminación es un tema crucial cuando se trata del uso de IA, porque el propósito principal de los algoritmos de aprendizaje automático es categorizar, clasificar y separar [...]. Como medio para afrontar la conformación de sesgos algorítmicos, los formantes de la “Rome Call for AI Ethics” apuestan por el cumplimiento de tres requisitos que permitan un alineamiento entre el avance tecnológico, el verdadero progreso del ser humano y el respeto por el planeta. El primero de ellos sería la inclusión de todos los seres humanos, sin que se produzca ningún tipo de discriminación. El segundo sería el mantenimiento en el centro de las acciones del bien de la humanidad y de cada ser humano. Y, por último, debe estar presente la compleja realidad de nuestro ecosistema y la forma en que se cuida y protege el planeta, con un enfoque altamente sostenible, quien incluye el uso de la IA para garantizar sistemas alimentarios sostenibles en el futuro». Igualmente, Hellman (2023: 201) considera que es evidente que «son muchos de los beneficios que entraña la IA. No obstante, debemos ser conscientes de que la utilización de algoritmos en el ámbito de la referida inteligencia artificial puede propiciar o alentar situaciones discriminatorias [...], por ello, debemos subrayar que, a la vista



Para llegar a este adecuado equilibrio, quizá sea necesario responsabilizar a las empresas para que demuestren que su tecnología no tiene efectos nocivos<sup>43</sup>. Cuando se crean sistemas que pueden generar tanta información a escala, debe probarse que no son dañinos. Por ahora, está sucediendo al revés, cada usuario tiene que demostrar los daños, e incluso, entonces, es bastante probable que no llegue a ninguna parte (Sánchez Díez, 2023a: 18). El ordenamiento jurídico ha conferido derechos fundamentales a la persona para proteger sus intereses, sin los cuales no podría desarrollarse de manera adecuada, según su biología y psicología. Debido a las nuevas formas de vulneración de estos derechos, como consecuencia de la evolución y los cambios sociales que están operando, el ordenamiento jurídico se debe esforzar por crear las barreras destinadas a defender a la persona de las nuevas intromisiones externas que puedan menoscabar sus derechos fundamentales (Núñez Zorrilla, 2019: 9). Por lo tanto, estos esfuerzos, aunque pasen por asignar responsabilidad a las corporaciones sobre sus tecnologías y efectos, y pueda esto entenderse por algunos como un freno al adecuado desarrollo de la IA, dejarían de ser «adecuados» y no se estaría atendiendo al doble compromiso de potenciar el desarrollo, uso y empleo de la sociedad, por un lado, y asegurando al mismo tiempo las garantías necesarias en el ejercicio de los derechos fundamentales, si finalmente se hubiera optado por imputar responsabilidad al propio sistema, por la vía de otorgarle personalidad jurídica. La ralentización

---

de los hechos, la normativa existente no es lo suficientemente eficaz. Resulta, por lo tanto, prioritario procurar nuevos mecanismos que tengan como objetivo asegurar y proteger de manera más contundente los derechos de las personas de cuyos datos se dispone. Éstos, desde nuestro punto de vista, deben fundamentalmente comprender un impulso de la regulación relativa a los derechos más básicos de los individuos, así como fomentar un mayor grado de transparencia».

<sup>43</sup> «Cuando la gente produce alimentos, tiene que demostrarnos que no son venenosos o que no contienen ingredientes perjudiciales para las personas. Aquí estamos hablando de información, la columna vertebral o de la sociedad actual. Como sucede con la forma de alimentarnos, impacta a todo lo que hacemos: afecta a nuestras decisiones y nuestra visión de la realidad; afecta a la democracia y cómo se comportan las personas» (véase Sánchez Díez, 2023a: 18). En esta misma línea, Beltrán de Heredia Ruiz (2023: 29) expone que «el avance de la ciencia se está desarrollando sin aparentas ser cuestionado. No solo se acepta resignadamente, como una consecuencia inevitable, sino que nos sentimos especialmente seducidos por ella: observamos con asombro cada avance y permanecemos expectantes a las nuevas versiones de los dispositivos. Esto explicaría por qué adoptamos una actitud colaborativa en aras a facilitar su expansión despreocupada. La pregunta es por qué somos insensibles a los efectos anteriormente descritos y permanecemos impasibles».

en la evolución de los sistemas de IA no necesariamente supone un límite, sino más bien un desarrollo sostenible y garante de los derechos fundamentales. Si las empresas, y todos los que intervienen en la cadena, hasta llegar al usuario de un sistema IA, supieran que son responsables de los daños que estos sistemas pudieran ocasionar, incluso cuando existe la posibilidad de aprendizaje autónomo, su manera de crear, publicitar, distribuir y utilizar sería muy diferente. En general, cuanto más sofisticados sean los sistemas de IA, más responsabilidad deberíamos exigir a sus diseñadores y programadores, para garantizar que cumplan principios legales y éticos. Las inteligencias artificiales no tienen, ni tendrán nunca, intencionalidad; son el reflejo de las intenciones y los sesgos de los equipos de programadores y entidades implicados en su implementación y su despliegue (López de Mántaras Badia, 2023b: 3).

Los agentes morales son las personas y no las máquinas. Por eso muchos expertos señalan la necesidad de regular su desarrollo e, incluso, de prohibir su uso en casos de riesgo extremo, como las armas letales. Pero, además de regular, es imprescindible educar a los ciudadanos, en particular a los políticos, sobre los beneficios y riesgos de las tecnologías inteligentes. Se necesitan futuros ciudadanos más y mejor informados, con más capacidad para evaluar los riesgos tecnológicos, con mucho más sentido crítico y capaces de hacer valer sus derechos. Este proceso de formación debe comenzar en la escuela y tener continuidad en la universidad. En particular, es necesario que los estudiantes de ciencia e ingeniería reciban una formación ética que les permita comprender mejor las implicaciones sociales de las tecnologías que desarrollarán. Solo si invertimos en educación conseguiremos una sociedad que pueda aprovechar las ventajas de las tecnologías inteligentes minimizando sus riesgos. Así, la inteligencia artificial servirá para realizar un gran paso en el progreso de la humanidad<sup>44</sup>.

### III. HACIA UN RÉGIMEN JURÍDICO ACORDE A LA CATEGORÍA JURÍDICA DE LA INTELIGENCIA ARTIFICIAL Y LA PROTECCIÓN DE LOS DERECHOS FUNDAMENTALES

Según se ha puesto de manifiesto (Ureña, 2019: 111), el primer paso antes de hablar de la «presunta» autonomía algorítmica, y el necesario equilibrio

---

<sup>44</sup> Véase López de Mántaras Badia (2023b: 3); en esta misma línea, Caramuto Martins (2013: 104): «Los ingenieros somos los que debemos hacer un ejercicio de ciencia ficción antes de desarrollar nuevos avances, sean cuales sean, y tratar de predecir el comportamiento de la sociedad ante dichos avances. La formación y el aprendizaje de los ingenieros no incluye actualmente objetivos en este aspecto, pero creo que sería algo que podría ayudar bastante a crear la sociedad de la información del futuro».

entre el desarrollo de la IA, así como una efectiva protección de los derechos fundamentales, es rechazar la denominada «falacia del homúnculo» (Balkin, 2015: 57). Es decir, la idea de creer que hay un agente tomando las decisiones autónomas del programa, una especie de persona que llega a conclusiones buenas o malas, cuando, en realidad, no existe tal persona. El algoritmo no es un agente al que se le pueda asignar responsabilidad ni hacer un reproche ético. Solo en el contexto de la sociedad humana tiene sentido el reproche al algoritmo. En este estado de cosas, podría decirse que las doctrinas jurídicas que impliquen un título de imputación de responsabilidad basado en algún tipo de reproche de conducta, como el dolo o la culpa en derecho penal, la negligencia en derecho civil o falla del servicio en derecho administrativo, no pueden predicarse respecto al algoritmo en sí mismo, sino del humano que lo utiliza (Ureña, 2019: 112).

En los procesos de *deep learning* serán los propios materiales descriptivos de la realidad desarrollados por los humanos<sup>45</sup> los que eduquen a los programas. De esta manera, el algoritmo aprenderá las diferencias de poder que se vean reflejadas en los materiales que se le provean o en los prejuicios del programador, por lo que, si estamos en una sociedad machista que produce datos machistas, el proceso de aprendizaje autónomo extraerá patrones machistas, por lo que es equivocado pensar que un algoritmo es machista, ya que machista es la interacción entre los humanos y la sociedad en la que opera el algoritmo<sup>46</sup>.

## 1. ALGUNAS CONSIDERACIONES SOBRE EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL APROBADO EL 13 DE MARZO DE 2024

Lo cierto es que, para garantizar la compatibilidad de la IA con los derechos fundamentales, es necesario que se establezcan marcos regulatorios sólidos y efectivos. Esto va a permitir que se identifiquen los posibles riesgos y limitaciones de la tecnología, al mismo tiempo que se fomenta su desarrollo y aplicación en beneficio de la sociedad en su conjunto (Castellanos Claramunt, 2023: 265). En este sentido, resulta esencial adoptar un enfoque multidisciplinar, que permita integrar consideraciones éticas, legales, sociales y tecnológicas en la regulación

<sup>45</sup> Nos estamos refiriendo a materiales como textos, fotos, vídeos o audios (Ureña, 2019: 106).

<sup>46</sup> Véase Ureña (2019: 107). Respecto a los datos de los que se alimentan los procesos de decisión autónoma, podría decirse que pueden obtenerse de tres maneras diferentes: 1) por el usuario que desarrolla sus propias bases de datos; 2) por la compara de esas bases de datos, y 3) por su extracción de las fuentes públicas (véase Levendowski, 2018: 583).

de la IA, con el fin de garantizar su compatibilidad con los derechos fundamentales y promover su uso responsable y sostenible (*ibid.*: 266). Las instituciones europeas han trabajado sin descanso durante los últimos años para asegurar que las nuevas tecnologías se basen en valores éticos y que la IA se desarrolle y aplique en un marco adecuado que promueva la innovación y respete los valores y derechos fundamentales de la Unión, en consonancia con la Carta de los Derechos Fundamentales de la Unión Europea<sup>47</sup>.

Como es bien sabido, en Europa, el pasado 13 de marzo de 2024, se ha aprobado el Reglamento de IA, también conocido como ley de IA<sup>48</sup>. Sus dos principales objetivos son, por un lado, garantizar que los sistemas de IA que se implementen en la Unión Europea sean seguros y respeten los derechos de los ciudadanos, y, por el otro, estimular la inversión e innovación en el ámbito de la inteligencia artificial en el marco europeo<sup>49</sup>. Sin embargo, aunque esta

---

<sup>47</sup> Véase Castellanos Claramunt (2023: 268), quien, además, expone los requisitos fundamentales que la Comisión Europea establece en su comunicado de 8 de abril de 2019 para asegurar la fiabilidad de la inteligencia artificial: 1) la necesidad de intervención y supervisión humanas para garantizar que los sistemas de IA promuevan sociedades justas; 2) la robustez y seguridad de los algoritmos para solucionar errores o inconsistencias que puedan surgir en todas las etapas del ciclo de vida de los sistemas de IA; 3) la privacidad y la gestión adecuadas de los datos personales; 4) la transparencia en el sentido de que los sistemas de IA deben ser trazables; 5) la diversidad, la no discriminación y la accesibilidad de los sistemas de IA para adaptarse a las capacidades, competencias y necesidades humanas; 6) la promoción del bienestar social y ambiental mediante sistemas de IA que fomenten la sostenibilidad y la responsabilidad ecológica, y 7) la rendición de cuentas mediante la implementación de mecanismos que permitan controlar los resultados de los sistemas de IA.

<sup>48</sup> El texto más reciente del reglamento se deriva de las modificaciones del propuesto en 2021, sobre todo por la aparición de la IA generativa en noviembre de 2022. Desde que, en diciembre de 2023, se llegara a un acuerdo sobre la regulación de la IA en el Consejo de la Unión Europea y el Parlamento Europeo, en marzo de 2024 este organismo ha aprobado el reglamento sobre la IA. Este ha sido respaldado por 523 votos a favor, 46 votos en contra y 49 abstenciones. Disponible en: <https://tinyurl.com/2knn3uee>.

<sup>49</sup> En este estado de cosas, resulta importante la definición que ofrece el reglamento, que considera que un sistema es inteligente «cuando se den las siguientes premisas: 1. Esté basado en máquinas. 2. Esté diseñado para funcionar con distintos niveles de autonomía. 3. Pueda demostrar capacidad de adaptación tras su despliegue. 4. Para objetivos explícitos o implícitos, infiere a partir de la entrada que recibe cómo generar salidas tales como predicciones, contenidos, recomendaciones o decisiones que puedan influir en entornos físicos o virtuales». En cuanto a los objetivos del reglamento, este busca: 1) mejorar el funcionamiento del mercado interior; 2) promover una inteligencia artificial centrada en el ser humano y digna de confianza,

ley da pasos positivos, en otras áreas, la legislación es débil y permite sistemas peligrosos en el ámbito de la migración, generando profundas desigualdades con sectores más desfavorecidos de la sociedad cuando el uso de la IA se realiza por parte de las autoridades policiales, de control migratorio y seguridad nacional<sup>50</sup>.

---

y 3) asegurar un alto nivel de protección de los derechos fundamentales. En su articulado recoge lo siguiente: a) normas armonizadas para introducir en el mercado, poner en servicio y utilizar sistemas de inteligencia artificial en la Unión; b) la prohibición de determinadas prácticas de IA; c) requisitos específicos para los sistemas de IA de alto riesgo y obligaciones para los operadores de dichos sistemas; d) normas de transparencia armonizadas para determinados sistemas de IA; e) normas armonizadas para comercializar modelos de IA de uso general; f) reglas sobre supervisión del mercado, sobre gobernanza de la vigilancia del mercado y sobre ejecución, y g) medidas de apoyo a la innovación, con especial atención a las pequeñas y medianas empresas, incluidas las de nueva creación. Por lo que se refiere a las principales obligaciones, el nuevo reglamento plantea nuevos esfuerzos que las entidades interesadas en estas tecnologías deben acometer para seguir utilizándolos. El reglamento presenta un enfoque basado en el riesgo. En consecuencia, impone un mayor número de obligaciones y exigencias a los sistemas de IA con niveles de riesgo más elevados. Asimismo, define cuatro niveles de riesgo asociados. 1) Sistemas de IA de riesgo inaceptable: a) representan una amenaza directa a la seguridad pública, la privacidad y los derechos fundamentales; b) su uso está prohibido, salvo en situaciones excepcionales como la utilización de sistemas de identificación biométrica remota por autoridades policiales en espacios públicos, con sujeción a salvaguardias. 2) Sistemas de IA de alto riesgo: a) pueden tener impacto relevante en los derechos fundamentales de los individuos; b) su uso está sujeto a ciertas obligaciones, como implantar sistemas adecuados de evaluación y mitigación de riesgos, registros de actividades, evaluaciones de impacto sobre los derechos fundamentales, medidas apropiadas de supervisión humana, etc. 3) Sistemas de IA de riesgo limitado: a) de propósito general; b) deben cumplir con ciertas obligaciones específicas de transparencia, como informar a los usuarios de que están interactuando con un sistema de IA o informar de que un contenido ha sido generado mediante IA. 4) De riesgo mínimo: a) no están regulados específicamente; b) los ciudadanos pueden decidir de forma libre sobre su uso (por ejemplo, videojuegos con IA o filtros de *spam*). Disponible en: <https://tinyurl.com/2knn3uee>.

<sup>50</sup> Así lo pone de manifiesto Mher Hakobyan, asesor de amnistía internacional para el trabajo de incidencia sobre la regulación de la IA. Disponible en: <https://tinyurl.com/3azxpsve>.

En esta misma línea, la coalición Protect Not Surveil expone que las prohibiciones sobre los sistemas de IA no se extienden al contexto migratorio. La legislación introduce prohibiciones (limitadas) para usos nocivos de la IA. Los legisladores de la UE se negaron a prohibir sistemas dañinos como los sistemas discriminatorios de evaluación de riesgos en la migración y el análisis predictivo cuando se utilizan para facilitar las

En cualquier caso, lo esencial es que la regulación de la IA debe garantizar la protección de los derechos humanos y evitar la discriminación, y no se puede ignorar que lo mismo que la ley es creada por el hombre, y establece exenciones que pueden ser discriminatorias y peligrosas, los algoritmos no son creados de manera autónoma, por lo que la debilidad de la legislación

---

devoluciones. Además, la prohibición del reconocimiento de emociones no se aplica en el contexto migratorio, por lo que se excluyen los casos documentados de IA en las fronteras. La lista de sistemas de alto riesgo no refleja los numerosos sistemas de inteligencia artificial utilizados en el contexto de la migración y que, en última instancia, no estarán sujetos a las obligaciones del presente reglamento. La lista excluye sistemas peligrosos como los sistemas de identificación biométrica, los escáneres de huellas dactilares o las herramientas de pronóstico utilizadas para predecir, interceptar y restringir la migración. La IA utilizada como parte de las bases de datos a gran escala de la UE en materia de migración, como Eurodac, el Sistema de Información de Schengen y ETIAS, no tendrá que cumplir con el reglamento hasta 2030. Exportación de tecnología de vigilancia dañina: la ley de IA no abordó cómo los sistemas de IA desarrollados por empresas con sede en la UE impactan en las personas fuera de la UE, a pesar de la evidencia existente de violaciones de derechos humanos facilitadas por tecnologías de vigilancia desarrolladas en la UE en terceros países (por ejemplo, China, territorios palestinos ocupados). Por lo tanto, no estará prohibido exportar un sistema prohibido en Europa fuera de la UE. Las autoridades antes mencionadas están explícitamente exentas de las normas y salvaguardias más importantes de la ley de IA.

1) Exenciones a la transparencia y salvaguardias de supervisión para las autoridades encargadas de hacer cumplir la ley. La ley introduce salvaguardias de transparencia que exigen a las autoridades públicas que utilizan sistemas de inteligencia artificial de alto riesgo registrar información sobre el sistema en una base de datos de acceso público. La ley de IA introduce una exención a este requisito para las autoridades encargadas de hacer cumplir la ley y de migración, inculcando el secreto para algunos de los usos más dañinos de la IA. Esto hará imposible que las personas afectadas, la sociedad civil y los periodistas sepan dónde están desplegados los sistemas de IA.

2) La exención sobre seguridad nacional permitirá a los Estados miembros eximirse de las reglas para cualquier actividad que consideren relevantes para la «seguridad nacional», en esencia, una exención general a las reglas de la ley de IA que, en teoría, podría usarse en cualquier asunto de migración, vigilancia y seguridad. Estas exenciones codifican efectivamente la impunidad para el uso irrestricto de tecnología de vigilancia, sentando un precedente peligroso para el uso de tecnología de vigilancia en el futuro. En efecto, los legisladores de la ley de IA han limitado enormemente el escrutinio crucial de las autoridades encargadas de hacer cumplir la ley y han permitido un uso cada vez mayor de una vigilancia racializada y discriminatoria. En primer lugar, estas lagunas jurídicas perjudicarán a los migrantes, a los racializados y a otras comunidades marginadas que ya son las más afectadas por los ataques y la vigilancia excesiva de las autoridades. Disponible en: <https://tinyurl.com/36kseenh>.

y de todos los algoritmos reside en los seres humanos que los construyen y las decisiones que toman respecto de los resultados (Castellanos Claramunt, 2023: 278).

Tal como se ha puesto de manifiesto (*op. cit.*), los problemas potenciales, entre los que se encuentra la discriminación, surgen a partir de las decisiones humanas involucradas en el proceso. Durante la fase de creación, esa discriminación puede ser fácilmente introducida, ya sea de forma intencional o voluntaria. Por lo tanto, se puede afirmar que un elemento esencial en la regulación de los algoritmos y su aprendizaje es la participación humana en su programación. En este estado de cosas, aunque los algoritmos sean precisos en su ejecución, eso no implica que sean verdaderos y respetuosos con los derechos fundamentales, ya que los datos utilizados para entrenarlos pueden estar sesgados debido a prejuicios del pasado<sup>51</sup>.

Por supuesto, ha de ser fácil y rápido establecer quién es el responsable de las consecuencias que puedan resultar de los actos de un sistema inteligente, siendo esenciales los mecanismos efectivos de reparación para las personas cuyos derechos fundamentales fueron infringidos por los sistemas automatizados de toma de decisiones. Sin embargo, lo verdaderamente importante es garantizar la efectiva protección de los derechos fundamentales, no su reparación. Ahí es donde reside la solidez del marco regulatorio. Por ello, la transparencia y explicabilidad algorítmica se convierten en una premisa esencial para lograr este objetivo<sup>52</sup>. En el Reglamento de IA, aprobado el 13 de marzo de 2024, ambos principios aparecen regulados en los arts. 13, 14 y 15 del referido texto normativo, aunque sus exigencias se antojan algo limitadas, tal y como ya sucedía en relación con el RGPD, en su art. 22 (Cotino Hueso, 2023: 56). La transparencia y explicabilidad son fundamentales para el conocimiento y la comprobación del buen funcionamiento del sistema de inteligencia artificial por los sujetos de la cadena de valor, tal es el caso de los usuarios del sistema, importadores o distribuidores, así como todos aquellos que tienen que verificar o comprobar este, tal es el caso de las autoridades y los evaluadores (*ibid.*: 21).

<sup>51</sup> Véase Kleinberg *et al.* (2018: 116). En este sentido, consideran que puede haber discriminación implícita si se utilizan registros de detenciones anteriores para predecir futuros delitos.

<sup>52</sup> También se hace referencia por parte de algunos autores a la importancia de considerar el control humano como un nuevo derecho y principio de la inteligencia artificial. Empresas como IBM, Microsoft, Telia Company e IA Latam se han sumado a esta discusión junto con países como Canadá, Japón, Estados Unidos y la comunidad política de la Unión Europea, a partir de declaraciones públicas y privadas en las que han manifestado la necesidad de considerar el control humano en la formulación y desarrollo de sistemas de inteligencia artificial (véase Sánchez Vásquez y Toro-Valencia, 2021: 5).

Por ello, tal y como ha expresado autorizada doctrina, es importante saber qué concretas transparencia e información de algoritmos son las debidas (*ibid.*: 17).

## 2. LA OPACIDAD DE LOS ALGORITMOS Y LA NECESIDAD DE CAMINAR HACIA LA TRANSPARENCIA Y LA EXPLICABILIDAD ALGORÍTMICA. UNA PREMISA INELUDIBLE EN LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL

Es urgente y necesario que aflore a la superficie legal el secretismo que esconden los sesgos de los datos absorbidos por el algoritmo. No pueden seguir siendo soberanos en la sombra de la sociedad digital, sobre todo en los ámbitos en los que podrían lesionarse derechos fundamentales (Castellanos Claramunt, 2023: 267). Tal y como se ha puesto de manifiesto (Ureña, 2019: 115), el desafío, en general, se ha planteado en términos de una «caja negra» (Pasquale, 2015: 1-19). Los procesos de decisión autónomos son opacos, en el sentido de que se conocen la entrada y el resultado, pero no se sabe de manera específica el proceso que llevó de uno al otro. Esta opacidad, aunada al hecho de que el proceso algorítmico reviste su resultado con un velo de objetividad e inevitabilidad, ha llamado a que se busque mayor transparencia, «abrir el algoritmo», para que se explique y los humanos puedan entenderlo<sup>53</sup>. Igualmente, la idea de

<sup>53</sup> En el Reglamento de Protección de Datos se regula lo que se ha venido denominando «el derecho a una explicación»: «Las personas tienen derecho a recibir una explicación de la decisión adoptada mediante la creación de perfiles y a recibir información significativa respecto a la lógica aplicada en ciertas decisiones automatizadas que afectan derechos fundamentales» (véase el considerando 71 del Reglamento de Protección de Datos). Disponible en: <https://tinyurl.com/4y86t4>.

El ejemplo más claro de ello se encuentra en el conocido caso *Loomis*. Durante la vista para decidir sobre su libertad condicional, el fiscal aportó un informe elaborado por un programa informático desarrollado por una empresa privada, según el cual el Sr. Loomis tenía un riesgo elevado de reincidencia y de cometer actos violentos, por lo que el juez impuso al señor Loomis una pena de prisión de seis años y otros cinco en régimen de libertad vigilada. La defensa del condenado recurrió la sentencia alegando que se había vulnerado el derecho a un proceso con todas las garantías, ya que no podía discutir los métodos utilizados por el programa informático, debido a que el algoritmo era secreto y solo lo conocía la empresa que lo había desarrollado. Si se observa esta cuestión dentro del ordenamiento jurídico español, no hay que perder de vista que el art. 24 de la CE reconoce a todas las personas el derecho a obtener la tutela efectiva de los jueces y tribunales en el ejercicio de sus derechos. No se trata solo de acceder a los órganos jurisdiccionales, sino del desarrollo posterior del proceso respetando los mínimos jurídicos básicos que comprende, tal y como ha manifestado



un control humano significativo es importante para garantizar la protección de los derechos fundamentales. Es decir, la creación de espacios algorítmicos para que un humano pueda revisar y retractar de manera efectiva la decisión del algoritmo, si ello fuera necesario (Ureña, 2019: 114).

Se trata de un cambio de paradigma que exige, ante el incumplimiento, pasar de un sistema de protección reactivo a un modelo preventivo y proactivo. Como una de las características de los sistemas de inteligencia artificial es la opacidad de los algoritmos para las personas, a las que les cuesta comprender su funcionamiento y la relevancia de sus aplicaciones, el cumplimiento del principio de transparencia resulta capital (Garriga Domínguez, 2023: 149). Tal y como se expresa en los arts. 13, 14 y 15 del Reglamento de Inteligencia Artificial, aprobado el pasado 13 de marzo, el responsable del tratamiento debe adoptar las medidas oportunas para facilitar al interesado toda información relevante relativa al tratamiento de sus datos personales incluidas la existencia de decisiones automatizadas y la elaboración de perfiles. Igualmente, debe facilitar toda la información significativa sobre la lógica aplicada, la importancia y las consecuencias previstas de dicho tratamiento para el interesado. Sin duda, el principio de transparencia obliga al responsable del tratamiento a garantizar el derecho de acceso a la información relativa a la existencia de decisiones automatizadas, incluida la elaboración de perfiles y, al menos en tales casos, información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento<sup>54</sup>.

Al hilo de estas directrices legislativas, resulta curioso descubrir, a través de la propia inteligencia artificial, cuál es su opinión acerca de la transparencia y explicabilidad de los algoritmos. En una entrevista realizada, a finales de diciembre de 2023, el propio ChatGPT, a la pregunta sobre la transparencia y accesibilidad de los algoritmos, respondía así: «Mis algoritmos no son transparentes ni accesibles al público. Soy un modelo de lenguaje desarrollado

---

en numerosas ocasiones el TC (véase Montero Claro y Castellanos Claramunt, 2022: 77). En esta misma línea, Goodman y Flaxman (2016: 51). Disponible en: <https://doi.org/10.1609/aimag.v38i3.2741>. El programa informático Compás utiliza más de cien factores, incluidos edad, sexo e historial criminal, para calificar a los acusados en una escala de 1 a 10: los calificados con puntuación de 1 a 4 son tratados como de bajo riesgo, mientras que los acusados con puntuación de 5 a 10 se tratan como de riesgo medio o alto. *State vs. Loomis* ha suscitado especial interés al ser la primera sentencia en Estados Unidos que permite que esos cálculos se hagan con un algoritmo cuyo funcionamiento es secreto (Freeman, 2016: 80; Martínez Garay, 2018: 494).

<sup>54</sup> Véase Cotino Hueso (2023: 33), quien considera que «se trata de un elemento clave que incide directamente en la calidad y robustez del sistema, así como respecto de la posibilidad de controlar sesgos, errores o posibles discriminaciones».

por OpenAI, específicamente el modelo GPT-3.5, y la arquitectura subyacente es bastante compleja. Mis algoritmos y la lógica interna de cómo genero respuestas no están disponibles para su inspección directa»<sup>55</sup>.

Actualmente, es prioritario y esencial que todos los esfuerzos se centren en adoptar medidas solventes para facilitar la transparencia y explicabilidad de los algoritmos, de manera que la protección de los derechos fundamentales sea realmente un hecho y no solo una posibilidad plasmada en los textos legales<sup>56</sup>. Hay quien considera que actualmente la Ley 1/2022, de 13 de abril, de transparencia y buen gobierno de la Comunitat Valenciana, ha de ser un referente para España y Europa, muy por encima de las escasas garantías normativas estatales y europeas existentes (Cotino Hueso, 2023: 55). Una vez sentadas las bases para una efectiva transparencia y explicabilidad, podrá atenderse con solvencia el régimen jurídico de responsabilidad por daños causados por sistemas de IA, aunque ese es ya otro tema<sup>57</sup>.

---

<sup>55</sup> En esta misma línea, ChatGPT continúa exponiendo que «la falta de transparencia en los algoritmos de inteligencia artificial ha sido un tema de discusión importante. La opacidad puede plantear desafíos en términos de explicabilidad y comprensión de cómo se toman las decisiones. Algunos esfuerzos se centran en el desarrollo de técnicas para hacer que los algoritmos sean más comprensibles y explicables, pero estas soluciones aún están en desarrollo. Es importante señalar que la falta de transparencia en mis algoritmos no significa necesariamente que estén ocultando información de manera maliciosa. La opacidad puede ser el resultado de la complejidad inherente a los modelos de lenguaje de gran escala como el mío. Sin embargo, las preocupaciones sobre la transparencia y la explicabilidad son válidas y deben abordarse a medida que la tecnología avanza».

<sup>56</sup> Es probable que el RD 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta del Reglamento del Parlamento Europeo y del Consejo, por el que se establecen normas armonizadas en materia de inteligencia artificial, haga tangible la transparencia y explicabilidad y permita que pueda verse el impacto que tiene en relación con la protección de los derechos fundamentales.

<sup>57</sup> Ciertamente, se trata de una puerta que ya se ha abierto. Sin embargo, quizá sea conveniente revisar el régimen de responsabilidad por culpa que se ha establecido en respuesta a los daños ocasionados como consecuencia de la implementación del uso de la IA en la vida cotidiana, y volver al sistema de responsabilidad objetiva o sin culpa, tal y como lo contemplaba la Propuesta de Reglamento del Parlamento Europeo de 2020 en relación con los SIA de alto riesgo. Es probable que esta hubiera sido una buena oportunidad para testear el comportamiento que ello hubiera generado en los creadores de algoritmos y toda la cadena de sujetos que se lucran o benefician con el desarrollo y la evolución de la IA, generando un adecuado equilibrio entre el desarrollo de la IA y la protección de los derechos fundamentales.

«La ciencia ficción se considera un ejercicio de fantasía, pero con el tiempo se ha comprobado que temas considerados como ciencia ficción se han convertido en realidad, el hombre ha llegado a la luna, y tiene robots y computadoras que trabajan para él, de modo que la imaginación del autor de la ciencia ficción acertó. Puede que realidad y ciencia ficción se acerquen, pero nunca llegarán a ser idénticas, ya que el trabajo del escritor es ir más allá. Cuanto más avance la ciencia, más lejos estará la base desde la que despega la imaginación del escritor de ciencia ficción. Hoy ya no puede considerarse a la ciencia ficción como algo para tomar en broma»<sup>58</sup>.

Isaac Asimov<sup>59</sup>.

### **Bibliografía**

- Adell, J. (2023). ¿Qué o quién puede hacer la escuela más inteligente? *elDiario.es*, 29-8-2023. Disponible en: <https://tinyurl.com/mr2r4st4>.
- Alaustrey, C. F. (2021). Estado de la cuestión de la inteligencia artificial y los sistemas de aprendizaje autónomo. *Revista Digital de Sociología del Sistema Tecnocientífico*, 11 (2), 182-195. Disponible en: [https://doi.org/10.24197/st.Extra\\_2.2021.182-195](https://doi.org/10.24197/st.Extra_2.2021.182-195).
- Álvarez Mellado, E. (2023). Pulpos, loros y sistemas conversacionales. *elDiario.es*, 26-8-2023. Disponible en: <https://tinyurl.com/ycxn67mf>.
- Aransay Alejandre, A. M. (2019). Antecedentes y propuestas para una regulación jurídica de los robots. En M. Barrio Andrés (dir.). *El derecho de los robots* (pp. 115-138). Madrid: Wolters Kluwer.

<sup>58</sup> En esta línea, es fácil fantasear con lo que pueda ser posible tras la demostración que Motos (2024) hizo tan solo hace unos días en su programa. En pantalla se mostró el impacto que la IA causa en los seres humanos cuando se trata de emular la voz de un ser querido ya fallecido. Todos los participantes, aun siendo conscientes de que no se trataba, en absoluto, de sus familiares, reaccionaban como si los tuvieran nuevamente a su lado. La capacidad del ser humano para empatizar con una máquina, que pueda mostrar ciertas características, denota que aún no se sabe hasta dónde puede llegar la «necesidad» de que se les dote de personalidad jurídica, como si de un ser humano se tratara, cuando, a día de hoy, ya se sabe que no es necesario, aunque sea posible. ¿De qué manera la realidad va a superar a la ficción en este caso? Ya se verá.

<sup>59</sup> A modo de recordatorio, las leyes de Asimov son la antesala «preceptiva» de toda regulación en materia de IA: 1) un robot no puede dañar a un ser humano ni, por inacción, permitir que un ser humano sufra daño; 2) un robot debe cumplir las órdenes de los seres humanos, excepto si dichas órdenes entran en conflicto con la primera ley, y 3) un robot debe proteger su propia existencia en la medida en que ello no entre en conflicto con la primera ley o la segunda (esto se da por sentado).

- Arrellano Toledo, W. (2023). La regulación de la Inteligencia Artificial sobre robótica. En W. Arrellano Toledo (dir.). *Derecho, Ética e Inteligencia Artificial* (pp. 224-245), Valencia: Tirant lo Blanch.
- Ataz López, J. (2020). Daños causados por las cosas: una nueva visión a raíz de la robótica y de la inteligencia artificial. En M. J. Herrador Guardia (dir.). *Derecho de daños* (pp. 317-375). Thomson Reuters Aranzadi.
- Balkin, J. (2015). The path of robotics law. *California Law Review*, 6, 45-61.
- Barrios Andrés, M. (2021). *Formación y evolución de los derechos digitales*. Santiago de Chile: Olejnik. Disponible en: <https://doi.org/10.33426/rcg/2021/110/1572>.
- Beltrán de Heredia Ruiz, I. (2023). *Inteligencia artificial y neuroderechos: la protección del yo inconsciente de la persona*. Navarra: Aranzadi.
- Beorlegui, C. (2023). Implicaciones filosóficas y perspectivas éticas de la Inteligencia Artificial. En R. Amo Usanos (dir.). *Inteligencia artificial y bioética* (pp. 1-210). Comillas: Universidad Pontificia.
- Boden, M. A. (2017). *Inteligencia artificial*. Madrid: Turner.
- Bolón-Canedo, V. (2023). Inteligencia Artificial vs. inteligencia humana: ¿Quién gana en el mundo actual? *elDiario.es*, 22-8-2023. Disponible en: <https://tinyurl.com/2j5u6x3w>.
- Caramuto Martins, G. (2013). *Influencia de la ciencia ficción en las tecnologías de la información y las comunicaciones y las telecomunicaciones*. Madrid: Escuela Universitaria de Ingeniería Técnica de Telecomunicación.
- Castellanos Claramunt, J. (2023). Sobre los desafíos constitucionales ante el avance de la inteligencia artificial. Una perspectiva nacional y comparada. *Revista de Derecho Político-UNED*, 118, 261-287. Disponible en: <https://doi.org/10.5944/rdp.118.2023.39105>.
- Chopra, D. (2020). *Metahumano*. Madrid: Gaia.
- Clarke, A. (1977). *Perfiles del futuro. Investigación sobre los límites de lo posible*. Barcelona: Luis de Caralt.
- Coca Payeras, M. (2023). Las iniciativas de la Unión Europea sobre Inteligencia Artificial: de la persona electrónica al difícil equilibrio entre la necesidad de impulsarla y evitar sus riesgos. *Revista de Derecho Civil*, 10 (2), 3-40.
- Copeland, J. (1993). *Inteligencia artificial*. Madrid: Alianza Universidad.
- Cotino Hueso, L. (2023). Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida. *Revista Española de la Transparencia*, 16, 17-63. Disponible en: <https://doi.org/10.51915/ret.272>.
- De Castro y Bravo, F. (1981). *La persona jurídica*. Madrid: Civitas.
- De la Quadra Salcedo Fernández del Castillo, T. (2018). Retos riesgos y oportunidades de la sociedad digital. En T. de la Quadra Salcedo y J. L. Piñar Mañas (dirs.). *Sociedad digital y derecho civil* (pp. 21-56). Madrid: BOE. Disponible en: <https://tinyurl.com/2w46t5tb>.
- Dennet, D. C. (2018). *La libertad de acción*. Barcelona: Gedisa.
- Díaz Calvarro, J. M. (2021). El robot como contribuyente. Reflexión sobre la personalidad jurídica digital. *Anuario de la Facultad de Derecho*, 37. Disponible en: <https://doi.org/10.17398/2695-7728.37.51>.

- Etxebarría Ecenarro, V. (2023). El desafío cuántico de la conciencia humana. *elDiario.es*, 21-8-2023. Disponible en: <https://tinyurl.com/kvzprvu2>.
- Fernández, C. B. y Cortés, I. (2018). Dos centenares de expertos europeos piden que no se reconozca personalidad jurídica a los robots. *LaLey.es*. Disponible en: <https://tinyurl.com/5n8e8bzt>.
- Freeman, K. (2016). Algorithmic injustice: how the Wisconsin Supreme Court Failed to protect due process rights in State vs. Loomis. *North Carolina Journal of Law and Technology*, 18, 75-106.
- García-Antón Palacios, E. (2023). El respeto a los derechos fundamentales desde la perspectiva ética de la Inteligencia Artificial. En A. M. Rocha Espíndola, D. Sansó-Rubert Pascual y N. Rodríguez dos Santos (coords.). *Inteligencia artificial y derecho* (pp. 170-194). Madrid: Dykinson. Disponible en: <https://doi.org/10.14679/2062>.
- Garriga Domínguez, A. (2023). Las exigencias de transparencia para los sistemas algorítmicos de recomendación, selección de contenidos y publicidad en línea, en el nuevo Reglamento Europeo de Servicios Digitales. *Revista Española de la Transparencia*, 17, 137-164. Disponible en: <https://doi.org/10.51915/ret.309>.
- Gavilán, I. G. R. (2023). Robots humanoides: desafíos éticos y normativos diferenciales. En W. Arrellano Toledo (dir.). *Derecho, ética e inteligencia artificial* (pp. 517-540). Valencia: Tirant lo Blanch.
- González de Patto, R. M. (2023). Inteligencia artificial y empleo. Análisis crítico del marco regulatorio europeo y español impulsado por el pilar europeo de derechos sociales. *Temas laborales: Revista andaluza de trabajo y bienestar social*, 168, 369-391.
- González Valenzuela, C. (2023). Retrasar el desarrollo de la inteligencia artificial 6 meses por su potencial peligro: ¿es buena o mala idea? *Computer*, 20-4-2023. Disponible en: <https://tinyurl.com/krba9hja>.
- Goodman, B. y Flaxman, S. (2016). European Union regulations on algorithmic decision-making and a right to explanation. *AI Magazine*, 38 (3), 50-57. Disponible en: <https://doi.org/10.1609/aimag.v38i3.2741>.
- Hellman, J. (2023). La violación 4.0 del Convenio Europeo de Derechos Humanos: la utilización de algoritmos discriminatorios en el ámbito de la inteligencia artificial. En A. M. Rocha Espíndola, D. Sansó-Rubert Pascual y N. Rodríguez dos Santos (coords.). *Inteligencia artificial y derecho* (pp. 201-222). Madrid: Dykinson. Disponible en: <https://doi.org/10.14679/2063>.
- Kleinberg, J., Ludwig, J., Mullainathan, S. y Sunstein, C. R. (2018). Discrimination in the age of algorithms. *Journal of Legal Analysis*, 10, 113-174. Disponible en: <https://doi.org/10.1093/jla/laz001>.
- Levendowski, A. (2018). How copyright law can fix artificial intelligence's implicit bias problema. *Washington Law Review*, 93, 579-630. Disponible en: <https://tinyurl.com/8yzvxh9t>.
- López de Mántaras Badia, R. (2021). La inteligencia artificial nunca será como la humana. *La Vanguardia*, 29-3-2021. Disponible en: <https://tinyurl.com/m7p486k9>.

- López de Mántaras Badia, R. (2023a). La inteligencia artificial aún no ha logrado nada que merezca el apelativo de inteligente. *elEconomista*, 18-4-2023. Disponible en: <https://tinyurl.com/bdw32c97>.
- López de Mántaras Badia, R. (2023b). Mitos y realidades de la inteligencia artificial. *Barcelona Metropolis*, julio de 2023. Disponible en: <https://tinyurl.com/2ezms7p3>.
- López de Mántaras, R. y Fernández Burgueño, P. (2019). Análisis de los retos, oportunidades y amenazas que conlleva la aplicación de la inteligencia artificial en el sector legal y en la sociedad en general. *elderecho.com*, 29-4-2019. Disponible en: <https://tinyurl.com/2m8h66fe>.
- Macanás, G. (2023). Bases para la personalidad jurídica de los entes no humanos. *Derecho Privado y Constitución*, 43, 11-52. Disponible en: <https://doi.org/10.18042/cepc/dpc.43.01>.
- Martín Casals, M. (2023). Las propuestas de la Unión Europea para regular la responsabilidad civil por los daños causados por los sistemas de inteligencia artificial. *InDret*, 3, 55-100. Disponible en: <https://doi.org/10.31009/InDret.2023.i3.02>.
- Martínez Garay, L. (2018). Peligrosidad y algoritmos *due process*: el caso *State vs. Loomis*. *Revista de Derecho Penal y Criminología*, 20, 485-502. Disponible en: <https://doi.org/10.5944/rdpc.20.2018.26484>.
- Montero Claro, M. D. y Castellanos Claramunt, J. (2022). Perspectiva constitucional de las garantías de aplicación de la inteligencia artificial: la ineludible protección de los derechos fundamentales. *Ius et Scientia*, 6 (2), 72-78. Disponible en: <https://doi.org/10.12795/IETSCIENTIA.2020.i02.06>.
- Motos, P. (2019). El robot humanoide Sofía, único en el mundo, llega al plató de «El Hormiguero» 3.0. *Antena 3*, 3-7-2019. Disponible en: <https://tinyurl.com/mwm4r8cr>.
- Motos, P. (2024). La experiencia real más emotiva de «El Hormiguero»: hablando con familiares fallecidos gracias a una Inteligencia Artificial. *Antena 3*, 31-1-2024. Disponible en: <https://tinyurl.com/28979wmw>.
- Nasit, S. (2017). Sophía, the robot wants to star a family. *Khaalej Times*, 23-11-2017. Disponible en: <https://tinyurl.com/5ah8ep5u>.
- Navarro Reyes, J. (2005). Cómo salir de la habitación china: conciencia e intencionalidad en las otras mentes. Debate sobre las antropologías. *Revista de Filosofía Thémata*, 35, 267-275.
- Núñez Zorrilla, M. C. (2019). *Inteligencia artificial y responsabilidad civil. Régimen jurídico de los daños causados por robots autónomos con inteligencia artificial*. Madrid: Reus.
- Ortego Ruiz, M. (2022). *La personalidad jurídica de los robots*. Valencia: Tirant lo Blanch.
- Otero Praga, M. (2023). ¿Puede la inteligencia artificial sustituir a la mente humana? Implicaciones de la inteligencia artificial en los derechos fundamentales y en la ética. *Anales de la Cátedra Francisco Suárez*, 57, 39-61. Disponible en: <https://doi.org/10.30827/acfs.v57i.24710>.

- Pasquale, F. (2015). *The black box society: the secret algorithms that control money and information*. Cambridge: Harvard University Press. Disponible en: <https://doi.org/10.4159/harvard.9780674736061>.
- Penrose, R. (1996). *La mente nueva del emperador. En torno a la cibernética, la mente y las leyes de la física*. Oxford: Oxford University Press.
- Perrigo, B. (2023). Open AI presionó a la Unión Europea para suavizar la regulación de la «I». *Time*, 20-6-2023. Disponible en: <https://tinyurl.com/y8fmzwzd>.
- Rebollo Delgado, L. (2023). *Inteligencia artificial y derechos fundamentales*. Madrid: Dykinson. Disponible en: <https://doi.org/10.2307/jj.5076311>.
- Riccio, T. (2021). Sophia robot: an emergent ethnography. *Threat Detection and Response*, 65, 42-77. Disponible en: <https://doi.org/10.1017/S1054204321000319>.
- Rodríguez de las Heras, T. (2023). No habrá régimen especial para la responsabilidad por daños causados por la inteligencia artificial. *Inese*, 3-6-2023. Disponible en: <https://tinyurl.com/2kvxcxrx>.
- Rogel Vide, C. (2018). Robots y personas. *Revista General de Legislación y Jurisprudencia*, 1, 79-90. Disponible en: <https://doi.org/10.30462/RGLJ-2018-01-04-646>.
- Sánchez-Castro Díaz-Guerra, C. (2023). *Estatus jurídico constitucional de los robots*. Valencia: Tirant lo Blanch.
- Sánchez del Campo, A. (2016). *Reflexiones de un replicante legal. Los retos jurídicos de la robótica y las tecnologías disruptivas*. Navarra: Aranzadi.
- Sánchez Díez, M. (2023a). Timnit Gebru: me gustaría que fuera responsabilidad de las empresas demostrar que su tecnología no tiene efectos nocivos. *elDiario.es*, 26-2-2024. Disponible en: <https://tinyurl.com/ypckkw2x>.
- Sánchez Díez, M. (2023b). Un ordenador racista está decidiendo tu futuro y tú no lo sabes. *elDiario.es*, 22-8-2023. Disponible en: <https://tinyurl.com/y4umhs8z>.
- Sánchez Vásquez, C. y Toro-Valencia, J. (2021). El derecho al control humano: una respuesta jurídica a la inteligencia artificial. *Revista Chilena de Derecho Tecnológico*, 2 (10), 1-12. Disponible en: <https://doi.org/10.5354/0719-2584.2021.58745>.
- Searle, J. (2001). *Mentes, cerebros y ciencia*. Madrid: Cátedra.
- Seth, A. (2023). *La creación del yo. Una nueva ciencia de la conciencia*. Madrid: Sexto Piso.
- Talavera Fernández, P. A. (2023). La superación del cuerpo y la desencarnación de la mente: imposibilidad ontológica de una máquina consciente. En J. Castellanos Claramunt (dir.). *Inteligencia artificial y democracia: garantías, límites constitucionales y perspectiva ética ante la transformación digital*. Barcelona: Atelier Constitucional. Disponible en: <https://tinyurl.com/3mpakmt5>.
- Taylor, H. (2016). ¿Podrías enamorarte de este robot? *CNBC*, 16-3-2016. Disponible en: <https://tinyurl.com/mr2f5nbr>.
- Turing, A. (1950). *Maquinaria computacional e inteligencia*. Santiago de Chile: Universidad de Chile.
- Ureña, R. (2019). Autoridad algorítmica: ¿cómo empezar a pensar la protección de los derechos humanos en la era del «big data»? *Latin American Law Review*, 2, 99-124, Disponible en: <https://doi.org/10.29263/lar02.2019.05>.