

ISSN: 0211-4526

PERSONA Y DERECHO

SEPARATA

2021* / 84
(ENERO-JUNIO)

MARTA ALBERT MÁRQUEZ POSTHUMANISMO, INTELIGENCIA ARTIFICIAL Y DERECHO

REVISTA SEMESTRAL DE LA FACULTAD DE DERECHO /
SERVICIO DE PUBLICACIONES DE LA UNIVERSIDAD
DE NAVARRA / PAMPLONA / ESPAÑA



Universidad
de Navarra

PERSONA Y DERECHO

REVISTA DE FUNDAMENTACIÓN DE LAS INSTITUCIONES JURÍDICAS Y DE DERECHOS HUMANOS
2021* / NÚMERO 84 (ENERO-JUNIO)

¿HUMANIDAD SIN FRONTERAS? TRANSHUMANISMO, POSTHUMANISMO Y DERECHO

Raúl MADRID / José Luis WIDOW / Luca VALERA (editores)

Prólogo 7-16

Vicente BELLVER CAPELLA

Transhumanismo, discurso transgénero y digitalismo:
¿exigencias de justicia o efectos del espíritu de abstracción? 17-53
[Transhumanism, transgender discourse, digitalism: Demands of justice or effects
of the spirit of abstraction?]

Giovanni SCARAFILE

Una ordinaria rarefacción. Brevi cenni sul nesso tra tecnologia e natura
umana nell'orizzonte del transumanesimo 55-73
[An ordinary rarefaction. Brief remarks on the link between technology and human nature
on the horizon of transhumanism]

Héctor VELÁZQUEZ FERNÁNDEZ

La revaloración del concepto de naturaleza humana en la sociedad
tecnológica. Una introducción a propósito del *human enhancement* 75-97
[The revaluation of the concept of human nature in our technological society. An
introduction in the context of the human enhancement proposal]

Jelson ROBERTO DE OLIVEIRA

Transformatio ad optimum: medicina estética e transhumanismo 99-118
[Transformatio ad optimum: aesthetic medicine and transhumanism]

Luca VALERA

Tres tesis sobre el transhumanismo 119-132
[Three theses concerning transhumanism]

Miguel ACOSTA

El camino hacia el transhumanismo: Génesis y evolución de un fenómeno
posmoderno a partir del «Mito del Hombre Nuevo» de Dalmacio Negro 133-156
[The way toward transhumanism: Genesis and evolution of a postmodern phenomenon
from the «Myth of the New Man» of Dalmacio Negro]

Cristina HERMIDA DEL LLANO Los riesgos del transhumanismo desde una perspectiva iusfilosófica [The risks of transhumanism from a philosophical-legal perspective]	157-181
Pablo LÓPEZ-SILVA Mejoras cognitivas y el problema de la inequidad: entre bioconservadurismo y bioliberalismo [Cognitive enhancement and the problem of inequality: between Bioconservatism and Bioliberalism]	183-206
Marta ALBERT MÁRQUEZ Posthumanismo, inteligencia artificial y Derecho [Posthumanism, Artificial Intelligence and Law]	207-230
Ángela VIVANCO MARTÍNEZ El hombre posthumano. <i>Singularity</i> y derechos fundamentales [The posthuman person. <i>Singularity</i> and fundamental rights]	231-255
José Luis WIDOW Un debate que falta. Sobre la necesidad de un adecuado concepto de libertad para fundamentar el <i>moral enhancement</i> [A Missing Debate. The need for a proper concept of freedom to ground moral enhancement]	257-285
Raúl MADRID Mejoras, mejoras humanas y mejoramiento moral biotecnológico [Enhancement, Human Enhancement and Moral Biotechnological Enhancement]	287-307
Rafael SANTA MARÍA D'ANGELO Transhumanismo, biotecnología y derechos humanos: diálogos, exigencias y necesidad de respuestas [Transhumanism, biotechnology and human rights: dialogues, requirements and the need for responses]	309-328
Jorge Nicolás LAFFERRIERE Transhumanismo, regulación de las tecnologías reproductivas y dignidad humana [Transhumanism, Reproductive Technologies Regulation and Human Dignity]	329-346
Pedro José IZQUIERDO La desigualdad del hombre «mejorado». El desafío transhumanista al orden jurídico liberal [The Inequality of the «Enhanced» Man The Transhumanist Challenge to the Liberal Juridical Order]	347-368
Normas para envío de originales	369

Posthumanismo, inteligencia artificial y Derecho

Posthumanism, Artificial Intelligence and Law

Marta ALBERT MÁRQUEZ

Universidad Rey Juan Carlos

marta.albert@urjc.es

<https://orcid.org/0000-0002-8805-0680>

RECIBIDO: 24/11/2020 / ACEPTADO: 03/05/2021

Resumen: El artículo se ocupa de analizar la contribución de la inteligencia artificial fuerte a la distopía posthumanista, detectando los presupuestos esenciales de la tesis que sitúa el futuro de la humanidad en la fusión de lo humano con una superinteligencia artificial. Finalmente, se extrapolan esas ideas fundamentales del posthumanismo al ámbito de la cultura jurídica, señalando algunos riesgos que de ellas se derivan para nuestra comprensión del derecho.

Palabras clave: Posthumanismo, Inteligencia artificial, Conciencia, Cuerpo, Derecho.

Abstract: The article is concerned with analyzing the contribution of strong artificial intelligence to post-humanist dystopia, detecting the essential assumptions of the thesis that places the future of humanity in the fusion of the human with an artificial super-intelligence. Finally, these fundamental ideas of post-humanism are extrapolated to the field of legal culture, pointing out some risks that derive from them for our understanding of law.

Keywords: Posthumanism, Artificial Intelligence, Consciousness, Body, Law.

I. INTRODUCCIÓN: PENSAR LO POSTHUMANO EN TIEMPOS DE PANDEMIA

La pandemia causada por la covid-19 obliga a la cuarentena de personas e invita a la de algunas ideas. Entre ellas ocupa un merecido puesto de honor el discurso transhumanista, cargado de promesas de invulnerabilidad encarnadas en el advenimiento de lo posthumano. Que un virus vaya ganando la partida a la ciencia, con más de 42,7 millones de contagios en todo el mundo y más de 1.1 millones de fallecidos¹ supone un acontecimiento que nos proporciona una mirada nueva sobre la fe depositada en la ciencia y en la técnica y en la capacidad de ambas para conducirnos a la erradicación del sufrimiento humano, a la victoria final sobre la muerte y, con ella, sobre las limitaciones propias de la especie.

Como señala Baltar, es muy probable que los discursos dominantes de la sociedad postmoderna, cuya volatilidad e inconsistencia ha puesto de mani-

¹ Datos obtenidos en <<https://www.rtve.es/noticias/20201025/mapa-mundial-del-coronavirus/1998143.shtml>> [30/10/2020].

fiesto la pandemia, fueran igual de inconsistentes y volátiles antes del mes de marzo de 2020². Pero lo cierto es que precisamente *ahora*, la perseverancia en predicar la apoteosis de la técnica resulta casi irrisoria. La «euforia tecnológica» bajo la que vivíamos, y que nos hacía pensar que la técnica podría librarnos de todo mal acechante³, se ha desinflado al ritmo vertiginoso al que han ido creciendo las cifras de muertos y enfermos por causa de la Covid-19.

Tal vez la pandemia constituya una oportunidad para replantearnos las expectativas de superación de nuestra propia humanidad desde otras muchas perspectivas, más allá de la cura de humildad científica que representa la situación que estamos viviendo prácticamente en todo el planeta. En todo caso, creo que merece la pena tomarse en serio las ideas que subyacen al pensamiento de lo posthumano. Al margen de lo quimérico de sus desarrollos prácticos, el posthumanismo se sostiene en tesis de carácter metafísico que suelen pasar desapercibidas o cuya verdadera naturaleza no siempre se hace explícita en el debate académico. Y la fuerza de esas ideas permanece bajo la incesante caída del tecno-optimismo colectivo⁴.

El objetivo de este trabajo consiste en mostrar, en términos necesariamente simples (espero que no simplistas), qué es el posthumanismo y qué papel juega en esta distopía contemporánea la inteligencia artificial. Trataremos de mostrar por qué los desarrollos esperables de la inteligencia artificial no nos conducirán hacia ninguna clase de «Singularidad» y, sobre todo, cuáles son los presupuestos metafísicos del sueño posthumano, cómo operan en nuestra cultura jurídica y qué consecuencias cabría esperar en este ámbito de su potencial desarrollo.

II. TRANSHUMANISMO Y POSTHUMANISMO: EL PAPEL DE LA INTELIGENCIA ARTIFICIAL

El movimiento transhumanista se define a sí mismo (empleo las palabras de Brostrom) como un «movimiento cultural, intelectual y científico que afirma la deuda moral para mejorar las capacidades físicas e intelectua-

² BALTAR, E., «El posthumanismo en la UCI de la realidad. Biología vs. Ideología», *Telos*, n. 114 (octubre, 2020), p. 86.

³ PONTIFICIA ACADEMIA PARA LA VIDA, *Pandemia y Fraternidad Universal. Nota sobre la emergencia Covid-19*, 30/03/2020. <http://www.academyforlife.va/content/dam/pav/documenti%20pdf/2020/Nota%20Covid19/Nota%20sobre%20la%20emergencia%20Covid-19_ESP_.pdf> [19/10/2020].

⁴ *Id.*, las conclusiones del Barómetro «Jóvenes y expectativa tecnológica» (septiembre de 2020), <<https://www.fad.es/notas-de-prensa/barometro-jovenes-y-expectativa-tecnologica/>> [19/10/2020].

les de la especie humana y para aplicar al hombre las nuevas tecnologías que puedan eliminar los aspectos indeseados y no necesarios de la condición humana: el sufrimiento, las enfermedades, la vejez e incluso su condición mortal»⁵.

En última instancia, el transhumanismo sostiene que es posible que tales mejoras nos conviertan (más bien, a nuestros descendientes) en «posthumanos», es decir, en seres más allá de lo humano, «seres que pueden tener, no solo una mejor esperanza de vida con calidad, sino también mayores capacidades intelectuales que ningún ser humano actual (y probablemente enteramente nuevas sensibilidades) así como la habilidad de controlar sus emociones»⁶.

Como afirma Albert Vidal «el transhumanismo-posthumanismo puede convertirse en la nueva utopía (o mejor dicho, distopía) del siglo XXI al prometer dejar atrás (como ya lo hace la eugenesia liberal) la discapacidad, lo imperfecto, el sufrimiento, la vulnerabilidad, la fragilidad, el envejecimiento, la caducidad, nuestra mortalidad, y en definitiva, todo lo que nos hace humanos, para esperar una especie de «salvación» mediante el advenimiento de una nueva condición posthumana»⁷.

Efectivamente, como subraya, entre otros, Birnbacher, la pretensión de introducir a la humanidad en un proceso hacia lo posthumano, intentando trascender ciertos límites inherentes a la condición humana tal como la conocemos, es un punto central del transhumanismo⁸. En ese mismo intento de superación de los límites de la condición humana, el transhumanismo aspira a hacer real una «libertad» absoluta, sin trabas⁹.

Max More expresó muy claramente ya en 1990, en el conocido artículo «Transhumanismo: hacia una filosofía futurista»¹⁰, el carácter episódico del transhumanismo, al definirlo como «un conjunto de filosofías que buscan guiarnos hacia una condición posthumana».

⁵ BOSTROM, N., *Intensive Seminar on Transhumanism*, Yale University Press, New Haven, 2003.

⁶ BOSTROM, N., «In defense of Posthuman dignity», *Bioethics*, vol. 19, n. 3 (2005), p. 203.

⁷ CORTINA, A., «Distopía posthumanista», en *Frontiere. Rivista di Geocultura* (08/03/2020), <<http://www.frontiere.eu/distopia-posthumanista/>> (19/10/2020).

⁸ BIRNBACHER, D., «Posthumanity, Transhumanism and Human Nature», en GORDIJN, B. y CHADWICK, R. (eds.), *Medical Enhancement and Posthumanity*, Springer, New York, 2008, p. 95.

⁹ Cfr. ALVIRA, R., «Qué nos hace humanos», en AA.VV., *Sobre Inteligencia artificial, ¿conciencia artificial? de Natalia López Moratalla*, Ideas y libros, Madrid, 2018, pp. 56 y ss.

¹⁰ MORE, M., «Transhumanism: towards a futurist Philosophy», *Extropy*, n. 6 (Summer 1990), p. 11.

Lo transhumano constituye una transición hacia lo verdaderamente posthumano. En lo posthumano ya no encontramos seres humanos más o menos modificados, sino, como su propio nombre indica, una realidad no humana. Lo posthumano implica el advenimiento de una nueva especie.

En el Manifiesto posthumanista¹¹ pueden encontrarse los postulados fundamentales del «credo» posthumano. No nos detendremos en todos ellos. Siguiendo el análisis de Torralba¹², partiremos de tres presupuestos básicos: Una fe ilimitada en el progreso; el carácter inexorable de la superación de la condición humana y la convicción de que estamos en el universo para ser superados por formas de vida superiores.

El posthumanismo implica, esencialmente, una reconsideración de lo humano y de su puesto en el universo, pero también una nueva consideración de la tecnología. En el paradigma de lo posthumano, no tiene sentido la vieja separación entre la naturaleza (lo humano) y lo artificial (la tecnología), ni siquiera la apología de la técnica que encontramos en el transhumanismo. Lo artificial pasa a formar parte de la definición de lo humano, o, por mejor decir, de lo posthumano. Como bien ha indicado Valera, el posthumanismo está destinado a superar la fractura entre lo que es técnica y lo que es hombre¹³.

Pero, ¿cómo surgirán esas formas de vida superiores destinadas a superar a los humanos?, ¿qué papel juega la inteligencia artificial en este proceso?

Como resulta obvio, la inteligencia artificial no es necesariamente «posthumana». Lo que sí es cierto es que se ha convertido en una pieza clave en el programa del posthumanismo. Quienes sostienen que la Singularidad está por venir, entienden que su llegada vendrá de la mano de los desarrollos futuros de la inteligencia artificial.

Así, el transhumanismo defiende que esa capacidad de incrementar nuestras capacidades físicas e intelectuales, así como de adquirir un mayor control sobre nuestros estados de ánimo, será adquirida no solo con la ayuda de las tecnologías que actualmente ya están de hecho implicadas en el mejoramiento humano, como la ingeniería genética o las tecnologías de la información,

¹¹ Publicado como apéndice del libro de PEPPERELL, R., *The Posthuman condition. Consciousness beyond the brain*, Intellect Books, Bristol, 2003.

¹² TORRALBA, F., «Deconstrucción del Posthumanismo. Los límites de la identidad humana», en SERRA, M. y CORTINA, A., *Humanidad infinita. Desafíos éticos de las tecnologías emergentes*, Tatiana Pérez de Guzmán el Bueno, Madrid, 2016.

¹³ VALERA, L., «Posthumanism: Beyond Humanism?», *Cuadernos de Bioética*, XXV, 3ª (2014), p. 487.

sino que también anticipa desarrollos futuros de otras, como la nanotecnología o la inteligencia artificial¹⁴.

¿Qué desarrollos futuros de la inteligencia artificial podrían llevarnos hacia lo posthumano?

III. LA «SINGULARIDAD TECNOLÓGICA»: EL ADVENIMIENTO DE UNA SUPERINTELIGENCIA GENERAL AUTÓNOMA

III.1. *Qué es la inteligencia artificial*

Antes de contestar a esta pregunta debemos poner brevemente en claro a qué nos referimos cuando hablamos de inteligencia artificial. La Inteligencia Artificial como forma de conocimiento es la «rama de las Ciencias de la Computación que estudia el software y hardware necesarios para simular el comportamiento y comprensión humanos»¹⁵. Como ciencia, su objetivo último es simular la inteligencia humana, creando robots de los que pueda predicarse tanto comprensión como conciencia (de esa comprensión y, en última instancia, de su propia existencia). Es lo que se conoce como inteligencia artificial «fuerte», empleando la terminología de Searle¹⁶.

Actualmente, el concepto de «inteligencia artificial» abarca un enorme campo de investigación, en el que se han obtenido resultados que van desde Siri hasta el armamento nuclear, pasando por ordenadores que nos ganan al Go! y coches que aparcen solos. Los ámbitos de investigación en IA están actualmente centrados en resolución general de problemas, creación de juegos para la competencia humana, análisis de imágenes y comprensión del lenguaje, entre otros¹⁷. Las máquinas nos superan en muchos aspectos (jugando al ajedrez, leyendo, resolviendo problemas matemáticos) pero se trata, en todo

¹⁴ *Ibid.* More cita «la neurociencia y la farmacología, las investigaciones sobre la extensión de la vida, la nanotecnología, la ultrainteligencia artificial, la exploración del suelo, combinado todo ello con una filosofía y un sistema de valores racionales», MORE, M., *op. cit.*, p. 6.

¹⁵ MALPICA VELASCO, J.A., *Inteligencia artificial y conciencia*, p. 1. <http://www3.uah.es/benito_fraile/ponencias/inteligencia-artificial.pdf> [19/10/2020].

¹⁶ SEARLE, J., «Minds, Brains and Programs», *Behavioral and Brain Sciences*, vol. 3, issue 3 (September, 1980).

¹⁷ RAFFIO, V., *Inteligencia Artificial. Entre la imaginación y la utopía*, Universitat Pompeu Fabra, Observatorio de Comunicación Científica, <<https://www.upf.edu/web/occ/4.-la-intel-ligencia-artificial>> [19/10/2020].

caso, de un empleo de la IA restringido a una determinada actividad concreta para la que la máquina ha sido previamente programada.

Esto puede expresarse de otra forma afirmando que, hoy día, los desarrollos de la IA permanecen, como mucho, en el ámbito de lo que se conoce como «inteligencia artificial débil», que, de nuevo con Searle, implica la clase de inteligencia artificial que permite diseñar y programar ordenadores de forma que realicen alguna tarea (concreta) que requiera «inteligencia». Pero no se trata, como en la concepción fuerte de la inteligencia artificial, de replicar (no simular) la inteligencia humana considerada en su conjunto (general) mediante una máquina¹⁸.

Una inteligencia artificial fuerte no es una *imitación* de la mente humana, significaría la posibilidad de generar una inteligencia humana en un soporte no biológico. Por tanto, la máquina inteligente en sentido fuerte sería capaz de comprender y de saber si comprende o no comprende, de tener conciencia de esa comprensión y de sí misma¹⁹. Algo así como HAL9000 en su *Odisea en el Espacio*. Ahora bien, ¿cómo se miden semejantes parámetros?

Debemos a Turing la primera formulación de qué se considera inteligencia en una máquina. Turing prefirió reducir el problema de si la máquina tenía inteligencia al de si la máquina podría imitar la inteligencia humana hasta el punto de confundir a un interlocutor humano, al que le sería imposible distinguir si hablaba con un ordenador o con una persona. La máquina que lograra tal cosa sería, según Turing, inteligente.

En la comunidad científica actual no hay consenso en torno a que ninguna máquina haya superado aún la prueba de Turing²⁰, pero tal cosa se considera esperable a corto plazo.

Sin embargo, de lo que no podemos estar seguros es de que la superación del test de Turing implique una verdadera inteligencia no biológica. La misma definición de inteligencia empleada es, a su vez, objeto de polémica. ¿Es válida la identificación entre inteligencia y la capacidad de imitación del pensamiento humano?

¹⁸ Vid., LÓPEZ MORATALLA, N., *Inteligencia Artificial. ¿Conciencia artificial?*, Digital Reason, Madrid, 2017, p. 80

¹⁹ ALFONSECA, M., «¿Basta la prueba de Turing para definir la inteligencia artificial?», *Scienta et Fides*, 2, 2 (2014).

²⁰ Hace algunos años se discutió si el chatbot Eugene Goostman, que logró «engañar» al 33% de los humanos en los que se hizo el test, podría estar cerca de ser «inteligente».

Searle respondió negativamente a esta cuestión, tratando de argumentar la diferencia entre que parezca que una máquina comprende y que comprenda efectivamente, a través del conocido ejemplo de la habitación china.

Con él trató de demostrar la ausencia absoluta de comprensión por parte de una máquina que podría responder con acierto a preguntas sobre una historia determinada, de modo que sus respuestas sugerirían que la había comprendido, cuando simplemente había ejecutado el algoritmo de Schank sin entender un ápice lo que contaba la historia²¹. La inteligencia artificial carece de semántica.

III.2. *Inteligencia artificial fuerte y Singularidad tecnológica*

Los desarrollos actuales de la inteligencia artificial están aún lejos de plantearnos un panorama posthumano, por más que estén emparentados con la mejora de las capacidades humanas que el transhumanismo entiende como un deber moral de los seres humanos²².

Pero, ¿y si desarrollos posteriores de la investigación nos pudieran acercar hacia un tipo de inteligencia artificial distinto, es decir, a una inteligencia artificial «general» o «fuerte»? Lo cierto es que tal cosa constituye el objetivo a largo plazo de un importante sector de investigadores en inteligencia artificial.

²¹ El argumento es el siguiente: «En primer lugar, imagina que las historias son contadas en chino y no en inglés –ciertamente un cambio que no es esencial– y que todas las operaciones del algoritmo de la computadora para este ejercicio concreto se suministran (en inglés) como un conjunto de instrucciones para manipular fichas con símbolos chinos en ellas. Searle se imagina a sí mismo haciendo todas las manipulaciones en el interior de una habitación cerrada. Las secuencias de símbolos que representan primero las historias, y luego las preguntas, se introducen en la habitación a través de una pequeña ranura. No se permite ninguna otra información del exterior. Finalmente, cuando se han completado todas las manipulaciones, la secuencia resultante se entrega a través de la ranura. Puesto que todas estas manipulaciones simplemente ejecutan el algoritmo de Schank, el resultado final será simplemente el equivalente chino de «sí» o «no», según sea el caso, con las que se responderá a una pregunta formulada en chino acerca de la historia también en chino. Searle deja en claro que él no entiende una sola palabra de chino, de modo que no tiene la más remota idea de lo que cuentan las historias», p. 25; *vid.*, PENROSE, R., *La mente nueva del Emperador En torno a la cibernética, la mente y las leyes de la física* (trad. José Javier García Sanz), Fondo de Cultura Económica, México, 1996, pp. 24-25.

²² Un reconocido ciborg es Neil Harbisson que ha creado la Cyborg Foundation en Barcelona. Según algunas previsiones, la «Cybor fusion» comenzará en torno a 2040, con la generalización de seres humanos portadores de distintos tipos de implantes electrónicos y alteraciones biológicas. Cfr. BARNATT, Ch., *Digital Genesis. The future of Computing, Robots and AI*, Wroclaw, Amazon Fulfillment Poland, Explaining Computers.com, 2017, p. 11.

Estaríamos entonces muy lejos de ganar o perder una partida de ajedrez. Estaríamos ante una superinteligencia que podría superarnos en cualquier tarea cognitiva. De hecho, para los partidarios de la inteligencia artificial fuerte, la máquina propiamente sería capaz de pensar. Y no solo eso, una superinteligencia general sería capaz de reproducirse de manera autónoma y dar lugar a sistemas cada vez más perfectos, hasta el punto de crear una forma de inteligencia capaz de marcar el inicio de una nueva etapa de la civilización, no necesariamente «humana».

Ya en 1965 Irving Good había llamado la atención sobre la «explosión» de la Inteligencia artificial. «La primera máquina ultrainteligente será la última invención que el ser humano necesite hacer, siempre y cuando la máquina sea lo suficientemente dócil para que nos diga cómo mantenerla bajo control»²³.

La obra de Good puede considerarse como un precedente del concepto de «Singularidad tecnológica». El término fue acuñado por Vinge en 1993. La Singularidad se correspondería con ese momento de «explosión» de la inteligencia artificial y supondría, según el propio Vinge, «el final de la era humana».

Pero sin duda el gran teórico de la Singularidad es Ray Kurzweil, uno de los más populares «gurús» del posthumanismo y fundador de la *Singularity University*. Según sus «predicciones», hacia el 2045 la inteligencia no biológica superará a la humana. El único modo de evitar la completa aniquilación de lo humano será fundirnos con la inteligencia no biológica. Nanorobots inteligentes estarán profundamente integrados en nuestros cuerpos, en nuestros cerebros. Como sabemos, esta integración nos permitirá superar la pobreza, nos otorgará una longevidad extendida, y, en definitiva, una vasta mejora de la inteligencia humana.

Arribaría entonces lo posthumano, esa nueva especie donde lo natural y lo artificial se vuelven indiscernibles. Pero aún no habría llegado la Singularidad. El posthumano es el precursor de la Singularidad, pero no debe confundirse con ella.

Según Kurzweil, la única manera de evitar nuestra desaparición, de evitar quedar superados a expensas de las máquinas superinteligentes es integrarnos

²³ GOOD, I., *Speculations Concerning the First Ultra-intelligent Machine*, 1965, <<http://acikistihbarat.com/dosyalar/artificial-intelligence-first-paper-on-intelligence-explosion-by-good-1964-acikistihbarat.pdf>> [19/10/2020]. *Vid.*, DVORSKY, G., «Why a superintelligent machine may be the last thing we ever invent», *Gizmodo* (2013), <<https://io9.gizmodo.com/star-wars-the-high-republics-enemies-are-much-more-cha-1845522125>> [19/10/2020].

en ellas. Se refiere fundamentalmente, como antes señalábamos, a una especie de «upload» de la inteligencia humana, que podría descargarse en el soporte no biológico de nuestra elección.

Para Bostrom, en cambio, nuestra «única salida» pasa por la programación. Es decir, sólo logrando una inteligencia artificial alineada con los fines y los valores humanos podríamos evitar la destrucción total de lo humano. El «alineamiento» de la inteligencia artificial con los valores humanos plantea dos problemas básicos, que aquí solo podemos dejar enunciados. El primero es de naturaleza técnica, y tiene que ver con cómo podríamos lograr el objetivo de construir una inteligencia artificial que compartiera nuestras prioridades y fuera sensible a nuestros valores e inquietudes. El segundo posee, en cambio, naturaleza normativa, ¿qué valores deberíamos enseñar a las máquinas?²⁴.

En definitiva, la Singularidad llegará el día en que la inteligencia no biológica tenga acceso a su propio diseño, y esté en condiciones de lograr su propia mejora en un ciclo continuo de rediseño. Cuando estos procesos autónomos lleven un ritmo tan rápido que la inteligencia humana sea incapaz de seguirlos, escapando ya totalmente a nuestro control, entonces habrá llegado la Singularidad²⁵.

La Singularidad tecnológica transformará la vida humana de manera irremediable. Según el propio Kurzweil, trascenderemos nuestros cerebros y nuestros cuerpos biológicos. «Aumentaremos el control sobre nuestros destinos, nuestra mortalidad estará en nuestras propias manos, podremos vivir tanto como queramos (que es un poco diferente a decir que viviremos para siempre), comprenderemos enteramente el pensamiento humano y expandiremos y aumentaremos enormemente su alcance»²⁶.

Tanto es así que, hacia el final del presente siglo la parte no biológica de nuestra inteligencia será «billones de billones» de veces más potente que la débil inteligencia humana producto de la biología.

Lo humano no dejará de existir, pero dejará de tener sentido la diferencia entre lo biológico y lo artificial. «No habrá distinción entre humano y máquina o entre realidad física y virtual»²⁷.

²⁴ GABRIEL, I., *Foundational philosophical questions on AI alignment*, Podcast, <<https://futureoflife.org/2020/09/03/iason-gabriel-on-foundational-philosophical-questions-in-ai-alignment/>> [19/10/2020].

²⁵ KURZWEIL, R., *Singularity* <<https://www.kurzweilai.net/singularity-q-a>> [19/10/2020].

²⁶ KURZWEIL, R., *La Singularidad está cerca*, Lolabooks, Berlin, 2019, p. 9.

²⁷ *Ibid.*, p. 10.

Tendremos la tecnología necesaria para «volcar» nuestra mente en una máquina superinteligente, que se irá automejorando hasta que de humano solo quede el origen de la inteligencia.

Kurzweil fundamenta su tesis de que la Singularidad está cerca en la conocida ley de Moore, que afirma que el número de componentes eléctricos de un circuito integrado se multiplicaría por dos cada año (si bien el propio Moore ha pensado que la Singularidad no es posible²⁸). En base a la ley de Moore, Kurzweil propone la ley de los «Rendimientos acelerados», sobre la que despliega su hipótesis de la posibilidad de un crecimiento exponencial de la inteligencia artificial. No niega Kurzweil la idea de que este crecimiento, como todos, entrará en algún momento en fase de meseta, pero está firmemente convencido de que la desaceleración tendrá lugar, en todo caso, una vez que la superinteligencia artificial se haya desarrollado hasta el punto de modificar para siempre nuestras vidas²⁹.

Otra de las razones que avalan el advenimiento de la Singularidad es el próximo cambio de paradigma tecnológico que traerá consigo la computación cuántica. Lo cierto es que la velocidad de cómputo registrada en las primeras pruebas realizadas con éxito en programación con *qubits* fue cien millones de veces superior a la de un procesador normal, pero aún es pronto para predecir el futuro de tales avances, y conocer su viabilidad real³⁰.

Sorprende el convencimiento con el que los singularistas expresan sus teorías sobre el futuro, sobre todo porque, más allá de las referencias a la ley de Moore o a los avances de la computación cuántica, ese convencimiento va acompañado de muy escasas referencias a la forma en la que será posible que ocurra todo esto.

De hecho, la creación de una inteligencia general, versátil, flexible y adaptable a los cambios, capaz de comprensión y de conciencia está aún muy lejos de la realidad. Cuando la revista *AI Magazine* publicó en 2006 el artículo titulado «Inteligencia artificial: los próximos veinticinco años»³¹ se puso de

²⁸ Cfr. DIÉGUEZ, A., «La Singularidad tecnológica y el desafío posthumano», *Pasajes: Revista de pensamiento contemporáneo. El inquietante siglo XXI*, n. 50 (2016), p. 157.

²⁹ KURZWEIL, R., *op. cit.*, p. 499.

³⁰ El procesador que utiliza IBM en Watson (D-Wave 2X) emplea esta tecnología, y efectivamente procesa algoritmos 100 millones de veces más rápido que un procesador ordinario. *Vid.*, ACOSTA, M., «AI Singularidad tecnológica y antropomorfismo», en AA.VV., *Sobre Inteligencia Artificial...*, *op. cit.*, p. 117.

³¹ STONE, M. y HIRSH, H., «Artificial Intelligence: The Next Twenty-Five Years», *AI Magazine*, 26(4), 85 (2005). <<https://doi.org/10.1609/aimag.v26i4.1852>> [19/10/2020].

manifiesto que ninguno de los veinticinco expertos consultados contemplaba la creación de una superinteligencia artificial ni de nada parecido al advenimiento de la Singularidad tecnológica.

La Comisión de Expertos de la UE que ha elaborado el documento «Directrices éticas para una IA fiable»³², cuyo objetivo es sentar las bases para la construcción de una inteligencia artificial centrada en el ser humano, tampoco se sitúa en ningún sentido en el escenario de la inteligencia artificial fuerte. Reconoce que su hipotético desarrollo es de hecho una preocupación para parte de la comunidad científica, pero que otra parte considerable de ella la considera pura ciencia ficción, por lo que sus recomendaciones van encaminadas al ámbito de la inteligencia artificial débil y, por tanto, no aborda ninguno de los problemas que se derivarían del empleo de una inteligencia artificial en sentido fuerte.

No es de extrañar que la Singularidad haya sido concebida como una especie de «caja negra». Según Acosta, es el recurso que se emplea para evitar decir «no sabemos cómo sucede esto», y, aún más, es la «varita mágica» con la que damos el salto de la física a la metafísica³³.

IV. ALGUNAS CONSIDERACIONES CRÍTICAS SOBRE LA SINGULARIDAD Y SU CERCANÍA

Como ya se ha señalado, los desarrollos actuales de la inteligencia artificial no autorizan a pensar que la Singularidad esté cerca, aunque la generación de una superinteligencia artificial esté en la mente de parte de la comunidad científica, representando la finalidad última de su investigación.

No entraremos en la dimensión técnica del problema, pero sí en el análisis de algunas preguntas que quedan en el aire, y que afectan a los presupuestos mismos de estos planteamientos. ¿Un progreso semejante de la computación cuántica daría lugar a la generación de una superinteligencia susceptible de introducirnos en la Singularidad?, ¿todos los problemas son computables, es decir, solubles mediante el empleo de algoritmos?, ¿la actividad mental puede

³² Dirección General de Redes de Comunicación, Contenido y Tecnologías (Comisión Europea), *Directrices éticas para una IA fiable*, 8 de noviembre de 2019, <<https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>> [19/10/2020].

³³ Como el propio autor subraya, google ha empezado a contratar filósofos. *Vid.*, ACOSTA, M., *op. cit.*, p. 117.

identificarse con algoritmos?, ¿podría, entonces, trasladarse de un soporte físico a otro, descargarse en un «chip», transmitirse sin más a otro cuerpo o a un sujeto no biológico, y con ella, a nosotros mismos?

Obviamente, tras esta filosofía de la inteligencia artificial fuerte subyace una idea de lo que significa el pensar humano mismo. Se presupone la identificación entre actividad mental y algoritmo (y, por tanto, el carácter exclusivamente mecánico del pensamiento) y, al mismo tiempo, la afirmación de una radical separación entre mente y cuerpo, es decir, la indiferencia del soporte físico en el que el algoritmo se encarna.

Como ha descrito Penrose, «según la IA fuerte, la diferencia entre el funcionamiento esencial del cerebro humano (incluyendo todas sus manifestaciones conscientes) y el de un termostato radica solo en que el primero posee una mucho mayor complicación (o quizá «mayor orden de estructura» o «propiedades auto-referentes», u otro atributo que pudiéramos asignar a un algoritmo). Y lo que es más importante, todas las cualidades mentales (pensamiento, sentimiento, inteligencia, comprensión, conciencia) deben ser consideradas, según este punto de vista, simplemente como aspectos de este funcionamiento complicado; es decir, son simplemente características del algoritmo que ejecuta el cerebro»³⁴.

Además, para las tesis que sostienen la inteligencia artificial fuerte, a este principio es necesario añadir el de la indiferencia del soporte material del algoritmo.

La sustancia mental, como habría dicho Descartes, no es materia. La estructura lógica del algoritmo es lo único representativo del «estado mental»: su encarnación física es indiferente. De ahí que Kurzweil cuente con la posibilidad de volcar nuestro cerebro en un soporte no biológico: porque ese soporte es del todo irrelevante para la existencia de la actividad mental. «Desde este punto de vista, una «conciencia humana» debe considerarse, en realidad, como un elemento de software, y su manifestación particular como ser humano material debe considerarse como la ejecución de ese software por el hardware de su cerebro y su cuerpo»³⁵.

Como antes señalábamos, la Singularidad opera aquí como la varita mágica que nos abre la puerta de la metafísica para que entremos en ella sin apenas aperecernos. Ambas afirmaciones distan mucho de poseer naturaleza

³⁴ PENROSE, R., *op. cit.*, p. 24.

³⁵ *Ibid.*, p. 31.

científica (y más aún de ser sostenibles metafísicamente). Por lo que respecta a la primera, parece obvio que la inteligencia humana no puede simularse en su totalidad recurriendo exclusivamente a procedimientos algorítmicos³⁶. Gran parte de nuestros procesos mentales son computacionales, pero no todos. De hecho, aunque resulte paradójico, todo apunta a que nuestra actividad cerebral no consciente sigue un patrón computacional, pero la conciencia opera de forma muy distinta, de modo que no puede describirse algorítmicamente³⁷.

No todo pensamiento es computacional, es decir, no toda forma de inteligencia puede resumirse en un conjunto de algoritmos. No lo son las intuiciones, ni tampoco el más elemental sentido común. A pesar de la asombrosa capacidad de aprendizaje en los ámbitos de la vida en los que cabe una «programación», las máquinas siguen sin aprender a resolver problemas muy básicos, como probablemente todos hemos pensado cuando se nos solicita rellenar un captcha³⁸.

El propio Turing, al plantear el problema de la parada, puso de relieve la existencia de problemas no solubles algorítmicamente³⁹, respondiendo así de forma negativa al dilema de la representación planteado por Hilbert (a saber, si existe un procedimiento mecánico capaz de resolver todo problema matemático).

En el ámbito de la propia matemática, se ha puesto de relieve cómo las intuiciones no pueden ser el resultado de operaciones puramente algorítmicas. La actividad mental a través de la cual los matemáticos llegan a sus juicios en torno a la verdad de las proposiciones matemáticas escapa a todo formalismo. Penrose ha tratado de demostrar, basándose en el teorema de la incompletitud de Gödel (en todo sistema formal hay una verdad no demostrable⁴⁰), que la verdad matemática está más allá de cualquier esquema formalista y que sería, por tanto, no computable en su totalidad, puesto que trasciende el formalismo⁴¹. En todo caso, las proposiciones en las que se vuelcan nuestros juicios

³⁶ *Ibid.* p. 362.

³⁷ *Ibid.* p. 366.

³⁸ Que viene a ser la inversa del test de Turing... Uno no puede evitar preguntarse cómo la infalible inteligencia artificial no puede hacer algo tan sencillo.

³⁹ El problema de la parada hace referencia a la imposibilidad de decidir que muestra la máquina ante determinados problemas.

⁴⁰ Me limito a extraer del teorema de Gödel la conclusión relevante para nuestro discurso. Obviamente se trata de una teoría infinitamente más compleja. Una descripción divulgativa de la misma puede encontrarse en GOLDSTEIN, R., *Incompleteness. The proof and paradox of Kurt Gödel*, Atlas Book, New York, 2005.

⁴¹ PENROSE, R., *op. cit.*, pp. 105-106.

conscientes sobre la validez de los algoritmos no pueden, a su vez, ser formulables algorítmicamente: esto parece ser indiscutible⁴².

Lo dicho puede verse más claramente si nos detenemos a considerar la inutilidad de las palabras para el pensamiento matemático (consideración que podríamos extender a otros ámbitos del pensamiento, también al ético, como veremos a continuación). Si el pensamiento no puede volcarse adecuadamente en palabras, no es tan difícil creer que pueda tener argumentos no algorítmicos⁴³. Las intuiciones esenciales de los conceptos matemáticos son simplemente «vistas» por los matemáticos. Lo mismo afirman los fenomenólogos respecto a las esencias de los valores que están «más allá del lenguaje» y que son directamente «dadas» en la reducción fenomenológica. El puntero de Husserl nos habla con suficiente claridad de esta experiencia prelingüística: el discurso puede tratar de ir mostrando esas verdades, solo al objeto de señalar hacia donde es preciso mirar... pero no puede mostrarlas o su naturaleza no se agota en el lenguaje⁴⁴. El hecho fenomenológico no es un hecho natural ni científico, su naturaleza es metafísica y por eso escapa a los límites del lenguaje y resulta refractario a cualquier intento de computación⁴⁵.

Por lo que respecta al segundo presupuesto de la inteligencia artificial fuerte, esto es, *la indiferencia del soporte del algoritmo*, se sostiene que la mente es la encarnación de un algoritmo suficientemente complejo activado por algunos objetos del mundo físico. No importa cuáles sean estos objetos reales. Señales nerviosas, corrientes eléctricas en cables, engranajes, poleas o tuberías.

Los desarrollos actuales de la ciencia parecen más bien sugerir lo contrario. Como afirma López de Mántaras, «el cuerpo da forma a la inteligencia y por lo tanto sin cuerpo no puede haber inteligencia»⁴⁶. De esta forma, para

⁴² *Ibid.*, p. 368.

⁴³ *Ibid.*, p. 378.

⁴⁴ «To be a pointer means to be silent» FRINGS, M., «Focusing on rarely seen complexities of Phenomenology», en SMITH, F.J., *Phenomenology in perspective*, Martinus Nijhoff, The Hague, 1970, p. 40.

⁴⁵ Para una exposición detallada del hecho fenomenológico, vid., SCHELER, M., «Lehre von den drei Tatsachen», en SCHELER, M. (ed.), *Gesammelte Werke*, Francke-Verlag, Bern, Bouvier Verlag, Muchen, Band X, 1957, pp. 449 y ss. Me he ocupado de la naturaleza de los hechos fenomenológicos en ALBERT, M., *Derecho y Valor: Una filosofía jurídica fenomenológica*, Encuentro, Madrid, 2004, pp. 67 y ss.

⁴⁶ LÓPEZ DE MÁNTARAS, R., «Algunas reflexiones sobre el presente y el futuro de la inteligencia artificial», en *Humanidad infinita. Desafíos éticos de las tecnologías emergentes*, op. cit., p. 97.

López Moratalla, el aprendizaje profundo que requiere la inteligencia artificial fuerte requiere necesariamente un cuerpo⁴⁷.

Dámaso apunta algo parecido desde el título mismo de una de sus obras más populares, *El error de Descartes*. Tal error consiste en haber defendido «la separación radical entre el cuerpo y la mente, entre el material del que está hecho el cuerpo, medible, dimensionable, operado mecánicamente, infinitamente divisible, por un lado, y la esencia de la mente, que no se puede medir, que no tiene dimensiones, es asimétrica, no divisible; la sugerencia de que el razonamiento, y el juicio moral, y el sufrimiento que proviene del dolor físico o de la conmoción emocional pueden existir separados del cuerpo. Más específicamente, que las operaciones más refinadas de la mente están separadas de la estructura y funcionamiento de un organismo biológico»⁴⁸.

Sin embargo, según la filosofía de la inteligencia artificial fuerte, la mente humana puede y debe ser «desprendida de su matriz biológica, para volcarla en un soporte más duradero y eficiente que el biológico»⁴⁹. Como Takeshi Kovacs, el asesino protagonista de *Altered Carbon*, uno debería poder cambiar de «funda» (así se denomina al cuerpo en la serie de Netflix) sin excesivo trauma. Lo esencial es la conservación del algoritmo al que se reduce nuestra identidad, vertido, en esta ocasión, en la pila que los posthumanos portan en la parte posterior de su cuello.

Fundas, hardware, algoritmos, software... como hace notar Juan Arana, se evita cuidadosamente hablar de materia y espíritu, pero el debate es básicamente el mismo. De nuevo, la cuestión no es enteramente científica. El punto en el que se sitúa la ciencia parece más bien reconocer que, sencillamente, los procesos cerebrales son los suficientemente complejos como para que sea imposible emplear las «arquitecturas informáticas como lecho de Procusto» para que esos procesos se acomoden a ellos⁵⁰.

⁴⁷ LÓPEZ MORATALLA, N., *op. cit.*, p. 115. López Moratalla explica como el cuerpo envía una retroalimentación al cerebro que este emplea para interpretar el mundo, y que constituye una información valiosísima. Algo tan prosaico como no poder fruncir adecuadamente el entrecejo por el uso del botox tiene como consecuencia un déficit en el procesamiento de las emociones negativas. La comprensión de frases tristes o que demuestren ira se vuelve muy lenta. Este fenómeno de ralentización no se da en el caso de que se trate de frases alegres, *ibid.*, pp. 145-146.

⁴⁸ DAMASO, A., *Descartes' error. Emotion, Reason and the Human Brain*, Putnam Sons, New York, 1994, p. 230

⁴⁹ ARANA, J., «Inteligencia natural e inteligencia artificial», *op. cit.*, p. 76,

⁵⁰ *Ibidem*.

La filosofía del siglo XX se ha esforzado, con resultados excepcionales, por poner de manifiesto el sentido último del cuerpo o, más bien, de nuestra condición corpórea. Desde Husserl en adelante, la filosofía fenomenológica de Marcel, Merleau-Ponty, Marías, Michel Henry o Ricoeur ha reaccionado contra el «tengo un cuerpo» afirmando la centralidad del cuerpo en la construcción de la identidad: «soy mi cuerpo». Mi cuerpo no es una funda, ni mi cerebro un ocasional hospedaje para mis pensamientos, innecesario o intercambiable.

V. ¿QUEDA ALGO DE HUMANO EN LO POSTHUMANO? LOS RIESGOS DE LA SUPERINTELIGENCIA GENERAL Y LOS RIESGOS DE LA IDEA DE UNA SUPERINTELIGENCIA ARTIFICIAL

Parece bastante evidente que, si la creación de una superinteligencia autónoma fuera posible, habría bastantes posibilidades de que no se tratara de una buena noticia para la humanidad en general, por más que los seres humanos estemos llamados a fundirnos con ella como método de supervivencia. ¿Quedaría realmente algo de humano en lo posthumano?

Incluso sin atribuir ninguna malevolencia a la máquina, hay que reconocer que su propio despliegue, la consecución de sus objetivos de manera ineludible (incluso en el supuesto de que esos objetivos estuvieran alineados con los nuestros) podría presentar obvios peligros para nuestra especie. Cualquier máquina, cumpliendo literalmente órdenes, incluso si son triviales, puede dar lugar a efectos desoladores⁵¹. Como señala Diéguez «no nos destruiría su malevolencia, sino su indiferencia»⁵². Y, obviamente, resulta ocioso subrayar las consecuencias de la inteligencia artificial, incluso débil, pero programada con un fin devastador para la Humanidad⁵³.

Por otra parte, cabe elucubrar cuáles serían las consecuencias, en términos jurídicos, del advenimiento de la nueva especie posthumana. La existencia

⁵¹ Es muy conocido el ejemplo de Stuart Russell sobre la máquina programada para servir cafés. Vid., RUSSELL, S., *Cómo la IA puede hacernos mejores personas*, Tedtalks, 2017, <https://www.ted.com/talks/stuart_russell_how_ai_might_make_us_better_people/transcript?language=es&hc_location=ufi> [19/10/2020].

⁵² DIÉGUEZ, A., *op. cit.*, p. 161.

⁵³ TEGMARK, M., *Benefits and risks of AI*, <<https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/>> [19/10/2020].

real de una inteligencia artificial fuerte, fundida con lo humano, pero rebasándolo con creces, podría traer consigo, como no resulta difícil imaginar, obvios problemas jurídicos. En parte relacionados con la fase de transición, donde podrían convivir con grave riesgo de discriminación lo humano con lo posthumano, y en parte con la configuración de una ordenación justa de las relaciones sociales en una comunidad de super-seres híbridos de naturaleza biológica y artificial a un tiempo.

¿Bajo qué paradigma podríamos elaborar la delimitación del derecho justo, una vez hubiéramos superado por completo la naturaleza humana? Esta última, en su sentido teleológico, había sido, como es notorio, el fundamento para la dinámica elaboración de las condiciones de posibilidad del derecho, sobre todo, entendido, como señala Finnis, en el sentido de a-justamiento⁵⁴.

Probablemente, la naturaleza posthumana impondría su propio paradigma para la construcción de algo parecido al derecho, es decir, de algún sistema de control social y ordenación de la conducta. Hasta qué punto podríamos o deberíamos llamarle derecho es una cuestión que rebasa la puramente terminológica. El derecho no tiene sentido en un horizonte donde no exista la debilidad ni la fragilidad, porque su razón de ser última radica, precisamente, en la protección de la vulnerabilidad humana.

En una sociedad de posthumanos invulnerables no existiría espacio para un orden jurídico en sentido estricto⁵⁵. Ni siquiera para la idea de los derechos humanos que, en última instancia, se fundamenta en la necesidad de protección jurídica del progresivo desenvolvimiento de la personalidad. Nada de esto tendría sentido una vez alcanzada la «perfección». ¿Tendría la naturaleza posthumana esas pulsiones hacia su propio bien que fundamentan los derechos? Es de suponer que no, puesto que a la vida como tarea (por expresarlo en términos orteguianos) sucedería la vida ya realizada, una vida, por lo mismo, sin fines ni metas.

Por pura economía del pensamiento, y teniendo en cuenta las obvias dificultades de todo tipo que debería superar el advenimiento de una inteligencia artificial fuerte y, a través de ella, de lo posthumano, prefiero centrarme no tanto en los riesgos que plantearía hipotéticamente la Singularidad tecnológica, sino en los riesgos que de hecho plantean las ideas que subyacen a la fe

⁵⁴ FINNIS, J., *Ley natural y derechos naturales*, Abeledo Perrot, Buenos Aires, Madrid, 1992, p. 235.

⁵⁵ Cfr. BALLESTEROS, J., «El derecho como no discriminación y no violencia», *Anuario de Filosofía del Derecho*, Madrid, 1974, t. XVII, pp. 159-165.

en la llegada de lo posthumano gracias a futuros desarrollos de la inteligencia artificial.

Porque quizá lo más relevante y lo más eficaz de los problemas que discutimos no sean sus implicaciones prácticas, sino las teóricas, y cómo estas, a su vez, pueden producir, y, de hecho, están produciendo, cambios significativos en nuestra autocomprensión, que se reflejan también, de algún modo, en la determinación de los rasgos fundamentales de nuestra cultura jurídica.

En realidad, como afirma Luca Valera⁵⁶, el verdadero poder del pensamiento posthumanista radica en su capacidad para redescubrir deseos y aspiraciones intrínsecamente humanos: la búsqueda de la eternidad y de la inmortalidad, el deseo de perfección, el sentimiento de soberanía sobre el cuerpo y esa tensión casi incesante que nos empuja a mudar nuestra condición de criaturas por la de demiurgos creadores.

Esas pulsiones (tan poco posthumanas) que inspiran el pensamiento posthumanista explican también su potencia, pese a sus escasas posibilidades de aplicación, y su capacidad para permear nuestra cultura y contribuir eficazmente a la construcción de una forma nueva de concebirnos no solo a nosotros mismos, sino también al derecho y al obrar jurídico.

VI. COROLARIO: IDEAS «POSTHUMANAS» PARA UN DERECHO DEMASIADO HUMANO

¿Cómo se proyectan estas ideas, que mantienen viva la esperanza distópica de lo posthumano, sobre el mundo jurídico? A continuación trataré de mostrar, a modo de corolario, que (1) algunos de los presupuestos de la idea de lo posthumano viven operantes en nuestra cultura jurídica, y que (2) estos presupuestos dan forma a aspiraciones tan antiguas como la humanidad misma, lo que acaso nos ayude a entender no solo la persistencia de la fe en lo posthumano, sino también la necesidad de apuntar precisamente a estas ideas «posthumanas» para neutralizarlas, impidiendo, de este modo, la construcción de un derecho «demasiado humano» (3).

En primer lugar, si trasladamos la posición de los partidarios de la inteligencia artificial fuerte en torno al pensamiento y su condición mecánica al ámbito de la reflexión sobre el derecho, vemos que todo apunta hacia una

⁵⁶ VALERA, L., *op. cit.*, pp. 489-490.

progresiva concepción del pensamiento jurídico como un pensamiento computacional, y esto es lo mismo que decir que todo ingrediente no mensurable de ese ejercicio creativo que constituye el pensar jurídico quedaría sistemáticamente fuera de lo asumible como pensamiento jurídico en sentido estricto.

1. No me es posible considerar aquí los problemas que el empleo de la inteligencia artificial «débil» pueda originar en el mundo jurídico, que son ya una realidad muy lejos de la ciencia ficción. El caso Loomis puso de manifiesto a un público amplio los problemas de la justicia algorítmica y la inevitabilidad del sesgo en la programación, y abrió un sustancioso debate sobre si realmente estaríamos dispuestos a ser juzgados por robots⁵⁷. Pero no estamos lidiando con las complicaciones que una inteligencia artificial débil pueda causar al derecho (aunque, sin duda, no son pocos ni irrelevantes).

Sí quisiera dejar apuntadas las implicaciones básicas de los presupuestos de la filosofía de la inteligencia artificial fuerte en nuestra concepción del Derecho, es decir, a la reducción del pensamiento jurídico a un mecanismo. El problema es tan viejo como la propia filosofía del derecho. Parecen resonar en nuestra memoria aquellas jocosas palabras de Pound: «pónganse los hechos en el orificio de entrada, empújese la palanca, y retírese la decisión predeterminada»⁵⁸.

La afirmación de que la praxis jurídica cabe en un algoritmo equivale a expulsar de ella los elementos en los que se constituye su propia especificidad, su carácter eminentemente creativo y no mecánico, la relevancia del sentido común, las intuiciones de los juristas y el sentimiento de la injusticia de los ciudadanos... Es más, equivale a la negación misma del derecho como ejercicio de razón práctica.

La estimativa jurídica no es computable, el puesto que juega nuestra intuición de lo justo en el obrar jurídico cotidiano no cabe dentro de un algoritmo y, sin embargo, es esencial a la «búsqueda del Derecho».

Antes examinábamos cómo las intuiciones matemáticas escapan a las posibilidades de computación y funcionan, a la vez, como el fundamento de la posibilidad de la matemática misma. La fenomenología del valor ha demostrado cómo esas intuiciones operan en todos los ámbitos del saber, y cómo

⁵⁷ Para más información sobre el caso puede consultarse <<https://confilegal.com/20180402-justicia-robotica-el-caso-loomis/>> [19/10/2020].

⁵⁸ POUND, R., «Law in Books and Law in Action», *American Law Review* (1910).

constituyen la condición de posibilidad de la construcción de la ciencia, precisamente en la medida en que escapan de la condición «científica», situándose más allá del lenguaje. Constituyen el objeto de lo que Max Scheler denominó «metafísica de los problemas límite de las ciencias»⁵⁹.

Nuestras intuiciones sobre la justicia delinean un espacio propio de cada cultura para la construcción del derecho. Un prejuicio infundado, fruto de la asunción, quizá inconsciente, de una tesis metafísica fuerte, como la de que solo existe el mecanismo, amenaza con expulsar de nuestra concepción del derecho todo aquello que le proporciona su sentido.

2. Si traducimos al lenguaje de los derechos la pretensión de libertad absoluta, de búsqueda de la eternidad y del señorío sobre uno mismo quizá entendamos mejor la contemporánea preeminencia del derecho a la autodeterminación personal, y la concepción que entiende el derecho mismo como un instrumento al servicio de la realización de los deseos narcisistas de cada uno de nosotros, entendidos como individuos aislados que buscan, cada uno en solitario, su propia autorrealización.

El transhumanismo y el posthumanismo representan formas diversas, aunque emparentadas, de hacer real una idea absoluta de la libertad, lo que, si se piensa bien, constituye una contradicción en los términos. Como afirma López Moratalla, «el ser humano es libre precisamente porque está condicionado. Ser libre significa la capacidad de liberación del encierro en las ataduras naturales»⁶⁰.

Desarrollos contemporáneos de la interpretación del derecho a la vida privada personal y familiar, en los términos de la Convención europea de derechos humanos, o del derecho a la intimidad y al libre desarrollo de la personalidad, en términos domésticos, tiene que ver con esta noción de la libertad narcisista que está llamada a propiciar su propio aniquilamiento. Nuevos derechos que se enraízan en deseos que aspiran a la protección del derecho aún desenvolviéndose al margen del bien común, es decir, al margen de razón alguna que avale su existencia entendiéndolos desde la consideración del ser humano como alteridad, como realidad abierta a otros y destinado a vivir una vida en común.

⁵⁹ SCHELER, M., «Philosophische Weltanschauung», en SCHELER, M. (ed.), *op. cit.*, Band IX, pp. 80 y ss.

⁶⁰ LÓPEZ MORATALLA, N., «Respuesta de Natalia López Moratalla a Rafael Alvira», en AA.VV., *Sobre Inteligencia artificial ¿conciencia artificial?*, *op. cit.*, p. 60.

El progresivo arrinconamiento de las libertades religiosa y de conciencia en nombre de la protección jurídica del desenvolvimiento del proyecto vital autónomamente concebido es otra muestra de la disposición de nuestro derecho a proteger de forma más eficaz al demiurgo que al hombre que reconoce una verdad y entiende que debe vivir conforme a ella. Como ha escrito Ballesteros «es una de las características constitutivas de nuestro ethos social, dominado por el pensamiento de la postmodernidad decadente: una voluntad «creadora» capaz de (re)configurar la realidad con el fin de satisfacer los deseos convirtiéndolos en derechos»⁶¹.

3. En tercer y último lugar, si trasladamos al mundo jurídico la indiferencia posthumana por el soporte físico del algoritmo en el que consistimos, obtendremos como resultado la progresiva irrelevancia de nuestro cuerpo, y, en particular, de su condición sexuada, en la construcción de la identidad jurídica postmoderna.

López Moratalla se refiere en su trabajo sobre inteligencia artificial a un proceso de triple emancipación del ser humano, que culminaría de alguna forma en las promesas del posthumanismo. Se trata de la separación del hombre de su origen, de su cuerpo sexuado y, en definitiva, de su propio cuerpo⁶². En realidad, son tres formas distintas de liberarnos de nosotros mismos, en una forma extrema de ejercicio de la libertad absoluta a la que nos acabamos de referir.

La separación del hombre de su origen, mediante la generalización de la anticoncepción y las técnicas de reproducción humana asistida, implica la inmersión de la generación humana en la lógica del dominio. Habermas ha mostrado suficientemente a dónde podrían llevarnos la mansa adopción de la eugenesia como filosofía vehicular de nuestro venir al mundo⁶³.

La separación del hombre de su condición sexuada (y del modelo binario que le es propio) es un segundo camino hacia la liberación de nuestra propia naturaleza, esta vez recorrido de la mano de las teorías del género, certificando que, como afirma Hadjadj, la explosión del sexo [durante la revolución sexual]

⁶¹ BALLESTEROS, J., «Exigencias de la dignidad humana en la biojurídica», en APARISI, A. y BALLESTEROS, J. (eds.), *Bioteología, dignidad y derecho: bases para un diálogo*, EUNSA, Pamplona, 2004, pp. 43-77.

⁶² LÓPEZ MORATALLA, N., *Inteligencia artificial, ¿conciencia artificial?*, op. cit., p. 40.

⁶³ HABERMAS, J., *El futuro de la naturaleza humana. ¿Hacia una eugenesia liberal?*, Paidós, Barcelona, 2012.

condujo a su desaparición»⁶⁴. Las recientes demandas de desaparición del sexo del estado civil de las personas nos hablan de esta «rebelión» frente a una hetero-conformación de la identidad sobre la base del ser-hombre o ser-mujer. El sexo «nos es asignado», en cambio, el género se presenta como un vehículo de conformación individual y autónoma de la propia identidad, liberada por fin de las ataduras naturales.

La ilusión de dejar atrás nuestra condición corpórea misma representa el último episodio de este itinerario emancipador. Dejo aquí solo apuntada una hipótesis: que nuestra progresiva incompetencia para entender el derecho mismo, basada en la opacidad de nuestra propia estructura relacional, trae causa del olvido del cuerpo, o, más bien, de la «absorción del cuerpo sexuado en la voluntad autónoma del sujeto, de forma que el cuerpo resulte comprensible solo como proyecto de autorrealización»⁶⁵. Si el cuerpo nos abre a la alteridad, si, como escribiera Mounier, es la primera frontera entre la persona y lo social⁶⁶, si existir en el cuerpo es existir relacionamente, puesto que el cuerpo constituye mi identidad y, a la vez, me lleva más allá de mí mismo hacia los demás... entonces «emanciparnos» del cuerpo significaría para nosotros volver imposible la misma idea derecho.

VII. REFERENCIAS BIBLIOGRÁFICAS

- ACOSTA, M., «AI, Singularidad tecnológica y antropomorfismo», en AA.VV., *Sobre «Inteligencia Artificial, ¿conciencia artificial?», de Natalia López Moratalla*, Ideas y libros, Madrid, 2018.
- ALBERT, M., *Derecho y Valor. Una filosofía jurídica fenomenológica*, Encuentro, Madrid, 2004.
- ALFONSECA, M., «¿Basta la prueba de Turing para definir la inteligencia artificial?», *Scientia et Fides*, 2, 2 (2014).
- ALVIRA, R., «Qué nos hace humanos», en AA.VV., *Sobre Inteligencia artificial, ¿conciencia artificial? de Natalia López Moratalla*, Ideas y libros, Madrid, 2018.
- BALLESTEROS, J., «El derecho como no discriminación y no violencia», *Anuario de Filosofía del Derecho*, t. XVII, Madrid, 1974, pp. 159-165.

⁶⁴ HADJADJ, F., *La profundidad de los sexos. Por una mística de la carne* (trad. F.J. Martínez y S. Moriel), Granada, 2010, pp. 60 y ss.

⁶⁵ GRANADOS, J., «Lenguaje del cuerpo, espacio sexuado y apertura al Creador: releyendo *Humanae Vitae*», *Anthropotes*, 34 (2018), p. 155.

⁶⁶ MOUNIER, E., *Le personalisme*, Presses Universitaires de France, Paris, 1950.

- BALLESTEROS, J., «Exigencias de la dignidad humana en la biojurídica», en APARISI, A. y BALLESTEROS, J. (eds.), *Bioteología, dignidad y derecho: bases para un diálogo*, EUNSA, Pamplona, 2004.
- BALTAR, E., «El posthumanismo en la UCI de la realidad. Biología vs. Ideología», *Telos*, n. 114 (octubre, 2020).
- BARNATT, Ch., *Digital Genesis. The future of Computing, Robots and AI*, Wroclaw, Amazon Fulfillment Poland, Explainingcomputers.com, 2017.
- BIRNBACHER, D., «Posthumanity, Transhumanism and Human Nature», en GORDIJN, B. y CHADWICK, R. (eds.), *Medical Enhancement and Posthumanity*, Springer, New York, 2008.
- BROSTOM, N., *Intensive Seminar on Transhumanism*, Yale University Press, New Haven, 2003.
- BROSTOM, N., «In defense of Posthuman dignity», *Bioethics*, vol. 19, n. 3 (2005).
- CORTINA, A., «Distopía posthumanista», en *Frontiere. Rivista di Geocultura* (08/03/2020).
- DAMASO, A., *Descartes' error. Emotion, Reason and the Human Brain*, Putnam Sons, New York, 1994.
- DVORSKY, G., «Why a superintelligent machine may be the last thing we ever invent», *Gizmodo*, 2013, <<https://io9.gizmodo.com/star-wars-the-high-republics-enemies-are-much-more-cha-1845522125>>
- FINNIS, J., *Ley natural y derechos naturales*, Abeledo Perrot, Buenos Aires-Madrid, 1992.
- FRINGS, M., «Focusing on rarely seen complexities of Phenomenology», en SMITH, F.J., *Phenomenology in perspective*, Martinus Nijhoff, The Hague, 1970.
- GABRIEL, I., *Foundational philosophical questions on AI alignment*, Podcast, <<https://futureoflife.org/2020/09/03/iason-gabriel-on-foundational-philosophical-questions-in-ai-alignment/>>
- GOLDSTEIN, R., *Incompleteness. The proof and paradox of Kurt Gödel*, Atlas Book, New York, 2005.
- GOOD, I., *Speculations Concerning the First Ultra-intelligent Machine*, 1965, <<http://acikistihbarat.com/dosyalar/artificial-intelligence-first-paper-on-intelligence-explosion-by-good-1964-acikistihbarat.pdf>>.
- GRANADOS, J., «Lenguaje del cuerpo, espacio sexuado y apertura al Creador: rele-yendo *Humanae Vitae*», *Anthropotes*, 34 (2018), pp. 143-170.
- HABERMAS, J., *El futuro de la naturaleza humana. ¿Hacia una eugenesia liberal?*, Paidós, Barcelona, 2012.
- HADJADJ, F., *La profundidad de los sexos. Por una mística de la carne* (trad. de F.J. Martínez y S. Moriel), Granada, 2010.
- KURZWEIL, R., *La Singularidad está cerca*, Lolabooks, Berlin, 2019.
- LÓPEZ DE MÁNTARAS, R., «Algunas reflexiones sobre el presente y el futuro de la inteligencia artificial», en SERRA, M. y CORTINA, A., *Humanidad infinita. Desafíos éticos de las tecnologías emergentes*, Fundación Tatiana Pérez de Guzmán el Bueno, Madrid, 2016.

- LÓPEZ MORATALLA, N., *Inteligencia Artificial. ¿Conciencia artificial?*, Digital Reasons, Madrid, 2017.
- LÓPEZ MORATALLA, N., «Respuesta de Natalia López Moratalla a Rafael Alvira», en AA.VV., *Sobre Inteligencia artificial ¿conciencia artificial? De Natalia López Moratalla*, Ideas y Libros, Madrid, 2018.
- MALPICA VELASCO, J.A., *Inteligencia artificial y conciencia*, <http://www3.uah.es/benito_fraile/ponencias/inteligencia-artificial.pdf>.
- MORE, M., «Transhumanism: towards a futurist Philosophy», *Extropy*, n. 6 (Summer 1990).
- MOUNIER, E., *Le personnalisme*, Presses Universitaires de France, Paris, 1950.
- PENROSE, R., *La mente nueva del Emperador. En torno a la cibernética, la mente y las leyes de la física* (trad. José Javier García Sanz), Fondo de Cultura Económica, México, 1996.
- PEPPERELL, R., *The Posthuman condition. Consciousness beyond the brain*, Intellect Books, Bristol, 2003.
- PONTIFICIA ACADEMIA PARA LA VIDA, *Pandemia y Fraternidad Universal. Nota sobre la emergencia Covid-19* (30/03/2020).
- POUND, R., «Law in Books and Law in Action», *American Law Review* (1910).
- RAFFIO, V., *Inteligencia Artificial. Entre la imaginación y la utopía*, Universitat Pompeu Fabra, Observatorio de Comunicación Científica.
- RUSSELL, S., *Cómo la IA puede hacernos mejores personas*, Tedtalks, 2017, <https://www.ted.com/talks/stuart_russell_how_ai_might_make_us_better_people/transcript?language=es&hc_location=ufi>.
- SCHELER, M., «Lehre von den drei Tatsachen», en SCHELER, M. (ed.), *Gesammelte Werke*, Francke- Verlag, Bern, Bouvier Verlag, Muchen, Band X, 1957.
- SCHELER, M., «Philosophische Weltanschauung», en FRINGS, M. (ed.), *Gesammelte Werke*, Francke- Verlag, Bern, Bouvier Verlag, Muchen, Band IX, 1976.
- SEARLE, J., «Minds, Brains and Programs», *Behavioral and Brain Sciences*, vol. 3, issue 3 (September, 1980).
- STONE, M., HIRSH, H., «Artificial Intelligence: The Next Twenty-Five Years», *AI Magazine*, 26(4), 85 (2005).
- TORRALBA, F., «Deconstrucción del Posthumanismo. Los límites de la identidad humana», en SERRA, M., CORTINA, A., *Humanidad infinita. Desafíos éticos de las tecnologías emergentes*, Tatiana Pérez de Guzmán el Bueno, Madrid, 2016.
- VALERA, L., «Posthumanism: Beyond Humanism?», *Cuadernos de Bioética*, XXV, 3ª (2014), pp. 481-491.