

TRABAJO FIN DE GRADO
GRADO EN FILOSOFÍA

LAS IMPLICACIONES FILOSÓFICAS DE
LA IA Y LA ROBÓTICA

Facultad de Humanidades. Sección de Filosofía
Curso 2016-2017

Alumna: Ruth Reyes Peraza

Tutor: José Manuel de Cózar Escalante

Índice

Introducción.....	5
1.Antecedentes.....	6
1.1. Humanismo y Posthumanismo.....	7
2. Estado actual.....	14
2.1. Una aproximación a la singularidad tecnológica.....	15
2.2. La convergencia tecnológica en la robótica.....	18
3. Discusión y posicionamiento.....	20
3.1. Responsabilidad.....	21
3.2. Autonomía de los robots e interacción con los seres humanos..	23
3.3. Tecnología y civilización.....	26
4. Conclusión y vías abiertas.....	34
Bibliografía.....	35

Introducción

Nick Bostrom, fundador del Future of Humanity Institute en la Universidad de Oxford, es uno de los representantes actuales más destacados de la reflexión sobre las Tecnologías Emergentes. “El filósofo del fin del mundo” (así calificado por The New Yorker Magazine) considera que cometemos un grave error al creer que todos nuestros problemas se verán solucionados por máquinas pensantes. Sin embargo, no adopta una postura totalmente pesimista ante los avances científicos y tecnológicos. Propone llevar a cabo una profunda reflexión, sin ánimo de rechazar las innovaciones tecnológicas. Tal es su confianza en las posibilidades de la ciencia que decidió fundar en 1998 junto a David Pearce la Asociación Mundial Transhumanista. Su convicción es que existe la posibilidad de una hibridación del ser humano con la tecnología y de potenciar así las capacidades humanas. Este filósofo sueco no considera tan descabellada la posibilidad de que llegado el momento la tecnología supere la inteligencia humana. Bostrom fundó también en 2004, junto a James Hughes, el Instituto para la Ética y las Tecnologías Emergentes, y en 2011 crea el programa Oxford Marton sobre los Impactos de la Tecnología Futurista. Es evidente su preocupación por los riesgos del futuro para la existencia humana.

En la actualidad, los expertos en ciencia y tecnología aseguran que la “Segunda Era de la Máquina” destruirá cinco millones de empleos en cinco años. Quizá pensemos en un primer momento que se trate únicamente de una cuestión política y económica. Sin embargo, la nueva propuesta por el parlamento europeo de hacer a los robots pagar impuestos implica un pensamiento filosófico sobre la sociedad que deseamos; cómo queremos vivir y trabajar, cómo interactuar con los robots y qué papel desempeñarán en la sociedad. Y es que dotar a los robots de una personalidad jurídica -ya que estos serán más autónomos, más inteligentes y serán capaces de aprender de su entorno- exige revisar en quién recaerá la responsabilidad civil de sus acciones. Esta cuestión ha suscitado un amplio y prolongado debate sobre si la ciencia y la tecnología implican un progreso para la humanidad, o más bien todo lo contrario. Antes del surgimiento de la computación moderna, el ser humano ha imaginado y reflexionado sobre los avances y riesgos que supone esta tecnología para la vida humana en el planeta. Sirva de ejemplo el escritor checo Karel Capek, quien en 1921 mencionó por primera vez en la pieza teatral *Los Robots Universales de Rossum* la palabra "robots",

describiéndolos como “unas criaturas mecánicas creadas para acabar con el trabajo manual del hombre”.

En 1959 tuvo lugar La Conferencia de Dartmouth, en el Dartmouth College, que reunió a prestigiosos científicos del campo de la computación (Oliver Gordon, Ray Solomonoff, Allen Newel, John McCarthy y Marvin Minsky entre otros). Uno de los acuerdos conseguidos en esta conferencia fue aceptar la propuesta de John McCarthy de acuñar el término de inteligencia artificial (IA) para la nueva disciplina que comenzaba a surgir. La inteligencia artificial provoca una enorme controversia, debido a la confusión y al desacuerdo que existe a la hora de concretar su definición. La IA tiene un importante valor práctico y, por tanto, puede ser comercializada. Ahora bien, se trata de una mercancía inusual porque no tiene forma material. En consecuencia, la robótica es considerada una extensión de la IA. Podemos incluso afirmar que constituye su apariencia material. Hoy día se está experimentando con robots capaces de negarse a seguir órdenes si las consideran peligrosas. Algunos robots ya logran aprender como si fueran niños mientras que otros "suben" sus conocimientos a internet para que los demás robots se beneficien de ellos. A este respecto, la robótica también tiene un valor cada vez mayor en el mercado económico.

El objetivo principal de este trabajo es debatir y analizar las principales implicaciones filosóficas del desarrollo y convergencia de la IA y la robótica en la actualidad. En el apartado de los antecedentes, comienzo con el esbozo de los principales hitos en el debate entre los humanistas reacios al avance técnico y los partidarios de estas nuevas tecnologías. Si bien para comprender la actual dimensión técnica del ser humano es preciso también rescatar el pensamiento que germina a finales del siglo XIX y continúa durante gran parte del siglo XX sobre el impacto de la ciencia y la tecnología en las personas y en la realidad social. Asimismo cabe preguntarse, ¿qué ha ocurrido para que se produzca la transición hacia otro paradigma? Se hará hincapié, por un lado, en la crítica al carácter legítimo de la ciencia y la tecnología. Para ello, recordaremos la postura de Friedrich Nietzsche, representante tanto de la crítica al humanismo como al saber científico. En tal sentido, parto de la sospecha nietzscheana de la domesticación del hombre a manos del conocimiento científico y humanístico. Ya desde la Ilustración se reafirma la exaltación y la confianza en la razón, y comienza a surgir la idea de que existe un cierto paralelismo entre el desarrollo racional y el progreso humano. Aunque mucho antes, Francis Bacon ya había planteado la idea de que el saber tiene que resolver los problemas fundamentales de la humanidad. Se produce entonces

un giro utilitarista y la razón se convierte en el instrumento para que el saber sea útil. De acuerdo con esto, en Nietzsche sitúo la denuncia de la idea central del cientificismo: eliminar al sujeto para conseguir la objetividad del saber, es decir, eliminar el mundo de la vida. Por esta razón, reivindico una reflexión acerca de las tecnologías emergentes de la IA y de la robótica, incorporando al sujeto, un sujeto que está en el mundo. También es un sujeto irracional, que desea, que tiene intereses y temores. En definitiva, se trata de un sujeto que no puede ser reducido a mera objetividad científica. Sin embargo, deseo aclarar que en este trabajo no se rechaza la posibilidad del conocimiento científico y tecnológico, pero sí se sugiere que es lícita la idea de que la ciencia y la tecnología están ancladas en el mundo de la vida y que directamente no implican un progreso.

Seguidamente, en el apartado referente al estado actual, muestro la convergencia de la robótica con otros campos de conocimiento y sus aplicaciones en la sociedad. Luego en la discusión y el posicionamiento continuo con la noticia anunciada en el parlamento europeo de que los robots deberán pagar impuestos. Esto constituirá el punto de partida para intentar comprender en qué nuevo paradigma nos estamos adentrando; si supondrá una revolución análoga a la revolución industrial; cómo afectará a nuestras vidas; y qué códigos éticos nuevos tendremos que formular. Atendiendo a estas consideraciones, la tesis a defender en este trabajo es la de que el progreso tecno-científico no implica un progreso directo en las condiciones materiales y sociales de vida. Vemos cómo hemos avanzado en ciertos aspectos tecnológicos y científicos, pero nuestros valores, costumbres y conciencias siguen demasiado estancadas en la crueldad y en la barbarie. De la miseria incluso se hace el mayor negocio, ya que hoy mismo innumerables personas huyen de la guerra en dirección a Europa. Los países más supuestamente civilizados hacen oídos sordos y pretenden "limpiar" su imagen, sus calles, y sus fronteras con objeto de evadir responsabilidades.

En la actualidad numerosos científicos están desarrollando espectacularmente la ciencia y la tecnología para la mejora y educación de la sociedad, tanto en términos teóricos como prácticos. Sin embargo, parece que no todos tenemos el mismo derecho y acceso a los avances científicos y tecnológicos. La ciencia no llega a ser verdaderamente pública, aunque se presente como tal, y así resulta que no decidimos acerca de las cuestiones fundamentales, como por ejemplo qué investigaciones se deben financiar y cuáles no, entre otras. Por lo tanto, en este trabajo planteo la cuestión de si las tecnologías emergentes de la robótica y de la IA suponen en realidad una mejora en la vida del planeta, o bien los problemas que

acarrear pueden acabar suponiendo lo contrario. No obstante los riesgos, considero que la solución no se encuentra en el rechazo absoluto de la tecnología (si es que éste pudiera producirse). ¿Por qué negar a las nuevas generaciones los avances en ciencia y tecnología? Aquí sitúo la filosofía como un pilar fundamental en el pensamiento crítico y en la conciencia del pasado, presente y futuro con respecto a la ciencia y las tecnologías en desarrollo. La filosofía es entendida desde este punto de vista como una herramienta crítica y esencial para comprender y reflexionar sobre el devenir histórico-tecnológico. Y es que, tal y como dijo Hölderlin, “mejorar el mundo no es ningún chiste”¹. En síntesis y citando a Bertrand Russell: “No hay más que un camino para el progreso en educación, como en todas las cosas humanas, y es el de la ciencia guiada por el amor. Sin ciencia, el amor es impotente; sin amor, la ciencia es destructiva”²

¹Friedrich Hölderlin, *Hiperión o el Eremita en Grecia*, prólogo de Jesús Munarriz. Libros Hiperión, 2005

²Bertrand Russerl, *Sobre educación*, prólogo de Jesús Mosterín. Editorial Espasa-Calpe, 1998

1. Antecedentes

Si no nos tomamos el tiempo indispensable para examinar el pasado, nos faltará la perspicacia necesaria para comprender el presente y dar rumbo al futuro, pues el pasado nunca nos abandona, y el futuro ya está aquí

LEWIS MUNFORD, *El mito de la máquina*

(*Técnica y evolución humana*)

Como primer acercamiento a las posturas humanistas y transhumanistas, o de la llamada *Humanidad Plus*³, expondré brevemente el paradigma ilustrado sobre la defensa de la razón. Se pretende abordar esta cuestión con el fin de comprender el carácter de la realidad social actual y qué decisiones se han de tomar para orientar el futuro de la especie y de la cultura humana.

El paradigma ilustrado reivindica un concepto de libertad que implica una transformación no solo de los métodos y fines del pensamiento, sino también de todas las estructuras sociales. Al respecto, Kant explica que la libertad y la dignidad humanas son valores absolutos cuyos intereses universales son exigibles a todo ser racional. Esta consideración de la libertad de acción también dio lugar a la organización política de la democracia liberal. Para lograr esta realidad social, el devenir teórico ilustrado se unió al desarrollo científico que se venía produciendo desde varios siglos atrás. Así, la ciencia y la ética fueron entendidas como un binomio y constituyeron el legado de la tradición ilustrada. En la cultura ilustrada se consideró a la razón como el único instrumento necesario para valorar y determinar lo que ha de hacerse. Por lo tanto, es la que fija los fines propios del hombre y los que han de regir también la acción científica.

³ El término transhumanismo fue acuñado en la década de los 50 por Julian Huxley

1.1. Humanismo y Posthumanismo

A partir del desarrollo del paradigma ilustrado surge un cambio de pensamiento histórico que influye enormemente en las corrientes del humanismo y el posthumanismo. El humanismo en sus inicios fue un movimiento intelectual que provocó la ruptura de las tradiciones escolásticas medievales y reivindicó lo racional como la cualidad más significativa de la naturaleza humana.⁴ El humanismo abarca una gran variedad de sistemas filosóficos e incluso varias escuelas de pensamiento religioso. A ello se suma su interés por los estudios clásicos y la filología grecorromana. Es evidente entonces que el concepto del humanismo reúne diferentes significados, si bien en este apartado no se atiende a la pluralidad del concepto del humanismo, sino que se hace una valoración crítica del movimiento humanístico en general.

Como primer exponente, podemos recordar la crítica al humanismo formulada por el filósofo Friedrich Nietzsche. Se basa principalmente en el estudio de la tragedia de la Grecia del siglo VI a. C. Nietzsche no solo se remitió a observar los hechos históricos del pueblo griego, sino que a su vez luchó por conseguir una nueva visión del significado de la tragedia griega. Se entendería la tragedia griega como el momento cultural que surgió del equilibrio entre lo apolíneo y lo dionisiaco. Esta idea tiene un alcance mucho más profundo que repercute en una revisión de la cultura occidental. Aquí podemos considerar que la filosofía de Nietzsche también se proyecta en el futuro abriendo un diálogo entre el mundo antiguo y el moderno. En la obra *El origen de la tragedia*⁵ Nietzsche parte del concepto de tragedia entendido como el equilibrio entre lo apolíneo y lo dionisiaco⁶, equilibrio que más tarde

⁴ Hay que tener en cuenta que el período de mayor esplendor del humanismo ocurre desde la Revolución Francesa hasta el final de la Segunda Guerra Mundial; luego, el progreso humanístico de la educación y la cultura es suplantado por los medios de comunicación y las tecnologías de la información.

⁵ Friedrich Nietzsche, *El origen de la tragedia*. Austral, 2013

⁶ Para Nietzsche Dioniso no es sólo un templo tal y como contemplaba en sus definiciones U. Von Wilamowitz, más bien entiende que el origen de la tragedia está en el culto de Dioniso unido al mito trágico y la música dionisiaca. En el capítulo 25 del *Nacimiento de la tragedia*, lo dionisiaco es comprendido como el fundamento de la existencia humana, es decir, la fuerza de la vida; y lo apolíneo es el aspecto ilusorio (aparente) mediante el que la experiencia de lo terrible y caótico se configura en el arte. Es importante comprender que para Nietzsche la existencia del mundo solo puede ser revelada como un fenómeno estético. El mundo es entendido entonces como una obra de arte y por lo tanto, la experiencia de la existencia es una experiencia estética similar a la ocurrida cuando contemplamos una obra de arte, o un bello paisaje.

entraría en declive a manos del surgimiento del racionalismo socrático. Prosiguiendo con este argumento y teniendo en cuenta la proyección del pensamiento nietzscheano con vistas al futuro, considero que esta reivindicación de la tragedia griega por parte de Nietzsche se debe a su intento por comprender, no solo la decadencia de la Grecia trágica, sino también la decadencia actual de la cultura occidental. La cultura romanizada vendría a señalar una decadencia con respecto a otra cultura mucho “más profunda y noble”⁷, que se corresponde con la Grecia arcaica y trágica. Asimismo, el planteamiento sugiere la separación del pensamiento nietzscheano de los grandes clásicos alemanes sobre el estudio de la Antigüedad y del método filológico convencional, denunciando la visión optimista del intelectualismo socrático. Con respecto a la pérdida del carácter trágico griego, Nietzsche considera que fue propiciado en un primer momento por Eurípides, antecesor de Sócrates, quien más adelante representó el espíritu ilustrado en su defensa de la razón sobre el instinto. También se sugiere la contraposición entre el optimismo intelectual y moral de Sócrates, y el pesimismo de la tragedia griega (el carácter antitético entre lo apolíneo y lo dionisiaco). Así, al producirse el rechazo del universo irracional del ser humano,- la dimensión simbólica de los rituales y mitos, la música, la poesía, la danza, el arte-, se muestra el triunfo del racionalismo socrático próximo al conocimiento científico.

Así vemos que Nietzsche parte del rechazo al racionalismo socrático para realizar una crítica al conocimiento científico y al humanismo. Según Nietzsche, el proyecto humanístico oculta la domesticación del hombre y por lo tanto, torna análogas la figura del sacerdote y del maestro, quienes en un principio se presentan como sus amigos. Sin embargo, considera que ocultan la imposición de un estado de control y de vigilancia, e incluso su domesticación y cría. En suma, la obra *El origen de la tragedia* gira en torno a la pregunta fundamental sobre qué es la vida. Nietzsche entendió el racionalismo como la negativa de la vida misma, esto es, el odio al mundo y el odio a la belleza. Con esto nos quiere decir que esta postura racionalista no acepta la inconmensurabilidad del caos, de la incertidumbre que encierra la vida y la muerte. Entonces, surge una voluntad de aniquilamiento de la vida y de la belleza. El pensamiento de Nietzsche plantea ya la cuestión del final del humanismo como utopía, e influye enormemente en el debate sobre la actualización del humanismo del siglo XX. Como apoyo a los argumentos de Nietzsche, Sloterdijk, en junio de 1997 en Basilea, provoca una enorme polémica retomando el concepto nietzscheano de domesticación y reclamando una

⁷ Ibid, p. 41

visión genético-técnica de la humanidad.⁸ En su obra, *Normas para el parque humano*, sugiere que existe una relación entre la histeria antitecnológica propia de los humanistas y la decadencia de la filosofía apoyada por la academia. Sloterdijk defiende la idea de que el humanismo y la filosofía son el resultado de una “secta de alfabetizados”⁹. De acuerdo con esto, los diferentes programas académicos de domesticación ocultan una lucha entre criadores, entre ideologías que respaldan diferentes procedimientos de crianza. Cuando Sloterdijk habla de “simples teorías de domesticación”, se refiere a que los hombres necesitan un amo, un criador; y cuando piensa en las “posibles técnicas de domesticación”, implica meditar sobre quién manda y cuál sería su finalidad.¹⁰ Por lo tanto, nos dice que la domesticación es el tema latente del humanismo, y sugiere incluso que “una lectura adecuada amansa”.¹¹

Para Sloterdijk la carta sobre el humanismo, escrita por Heidegger en 1946, no deja de ser “una carta voluminosa para los amigos”. De esta manera, entiende la crítica de Heidegger al humanismo como otra continuación del proyecto humanista. La obra de Heidegger *ser y tiempo* está pensada en contra del humanismo. En esta obra Heidegger subraya el sentido

⁸ Para comprender qué es el hombre según Sloterdijk, es necesario ir más allá del humanismo mediante un ejercicio genealógico, que fue iniciado por Nietzsche en su *Genealogía de la moral*. Sin embargo, la genealogía que propone Sloterdijk también recibe una enorme influencia del sociólogo y filósofo Arnold Gehlen en sus escritos *El hombre en la era de la tecnología* 1957, donde desarrolló su tesis del hombre como ser “deficitario”. Esto significa que el hombre, ante la imposibilidad orgánica de adaptarse al medio ambiente, se ve obligado a crear un medio ambiente artificial que le permite producirse a sí mismo con independencia del mundo orgánico. Según Gehlen la *Techné* no son herramientas que el hombre fabrica, sino más bien la técnica es el resultado de la inteligencia práctica del hombre que le permite disponer del entorno y someterlo a las necesidades vitales. Con esto no se quiere decir que el hombre haga uso de la técnica, sino que el hombre es en sí mismo un animal técnico. Por lo tanto, para Sloterdijk el mundo primitivo de la tribu no era naturaleza pura, sino un entorno artificialmente producido, y la función de este mundo artificial no es otra que la crianza de seres humanos.

⁹ Peter Sloterdijk, *Normas para el parque humano: una respuesta a la carta sobre el Humanismo*. Ediciones Siruela, 2008, p 24. Sloterdijk utiliza como ejemplo de antropotécnica al humanismo, ya que éste consiste en el compromiso por parte de algunos hombres de rescatar a otros de la barbarie. El humanismo es el intento de crear una minoría de alfabetizados que sean capaces de inhibir el despliegue de la cultura analfabeta de las masas.

¹⁰ Sloterdijk se refiere a la antropogénesis como el proceso a través del cual los hombres generan una naturaleza que les permite autorproducirse y autocriarse.

¹¹ Ibid, p.32. Según Sloterdijk el proyecto humanista que aboga por la educación y la cultura oculta la disputa por la antropogénesis, es decir, la lucha por conseguir un derecho tutorial y procreador sobre el hombre.

profundo del lenguaje y por lo tanto, replantea la pregunta por el ser, se refiere, a la pregunta por el ser del *Da-sein*. El sustantivo *Da-sein* significa entre otras cosas la vida, la existencia humana, que es a lo que Heidegger apunta. Además, Dasein, por aquello que “sein” en alemán es “ser”, es un modo de ser; “da” lo especifica en cuanto tal modo de ser, de esta forma, el verbo Dasein se corresponde con el existir, aun mejor, con el haber (lo que hay); pero sobre todo el estar, aquí, allí, es decir, en un lugar y momento determinado.¹² De acuerdo con esto, el existir al que Heidegger se refiere es siempre el propio, el de uno mismo, y aquello en lo que uno mismo está. Así, “yo estoy” es siempre por definición “estar aquí”, ser en una determinada situación. Heidegger habla del Dasein como ser-en-el-mundo, de manera que en toda interpretación del mundo está comprendida nuestra existencia y viceversa. El Dasein como ser-en-el-mundo siempre es un mundo ya conocido, ya interpretado de este o aquel modo. Ese hecho peculiar, que el mundo al que vengo, en el que crezco, está para mí ahí en un determinado ser-interpretado, implica que todo mirar, toda perspectiva está determinada en un sentido concreto. Con todo, ¿qué quiere decir Heidegger cuando reafirma que no se ha alcanzado este sentido del ser de la existencia?¹³ Heidegger entiende que la pregunta por la esencia del hombre no tomaría un rumbo acertado hasta que no se distancie de la metafísica europea, que defiende al hombre como un ser racional. Contra esto se sitúa el análisis existencial-ontológico de Heidegger. Para él la esencia del hombre no puede ser expresada desde una perspectiva zoológica o biológica. En este sentido, Sloterdijk critica a Heidegger porque considera ridícula su forma de hablar de la existencia humana, y en último término, porque los hombres son colocados “ahí” en su modo de ser, como guardianes del ser. Así nos dice que “Heidegger pretende un hombre mucho más sumiso que el mero buen lector”¹⁴. Prosiguiendo con la crítica de Sloterdijk a la comprensión

¹² Martin Heidegger, *Ontología: Hermenéutica de la facticidad*. Alianza, 2000

¹³ Heidegger se pregunta también acerca del límite que alcanzará la técnica con su capacidad para destruir la civilización. Heidegger en su texto, *la pregunta por la técnica*, se propone preguntar sobre la esencia de la técnica, de manera que con preguntar quiere abrir un camino para pensar cuál es la relación entre la esencia de la técnica y lo que corresponde a sus limitaciones. Para Heidegger es importante no confundir la técnica con la esencia de la técnica, ya que ésta última, “no es, en absoluto, algo técnico”. Con esto Heidegger, advierte que, por una parte se da una concepción instrumental y antropológica de la técnica, pero dicha concepción de la técnica, no nos lleva a alcanzar su verdad, es decir, conseguir desocultar la esencia de la técnica. Al respecto dice: “solo allí donde acontece tal desocultar, acontece lo verdadero”. En este sentido, Heidegger quiere llegar a las entrañas mismas del hombre y su pensamiento está articulado en torno a la crítica dirigida a occidente. Así, el hombre ha olvidado el ser, y en su lugar ha tomado respuestas infundadas desde fuera, o bien impuestas.

¹⁴Ibid., p. 47.

heideggeriana de la existencia, nos dice que Heidegger no sólo presupone de nuevo la palabra “humanismo”, sino que también señala que se ha perdido su sentido. Al respecto, Sloterdijk plantea la cuestión de si es necesario conservarla. Con todo, aunque Heidegger se plantee la pregunta por el fundamento de la domesticación; Sloterdijk le critica a Heidegger que en su concepción del hombre como ser-ahí y ser-en-el-mundo se sitúa en contra de toda antropología técnica, es decir, de la revolución antropogenética.¹⁵ Por eso, según Sloterdijk, el nuevo espacio de pensamiento posthumanista sobrepasa al pensamiento heideggeriano.¹⁶

Las controvertidas declaraciones de Sloterdijk implican ciertas consecuencias con respecto a la base del humanismo. Todo el humanismo construido sobre el ideal alfabetizado del hombre se derrumba y obliga a replantear el proceso educativo desde la perspectiva de la cría y domesticación, así como desde las nuevas tecnologías. Sloterdijk, al situarse en contra de las bases del humanismo, abre un nuevo horizonte técnico que hay que tener en cuenta. Es interesante puntualizar que Sloterdijk considera que la domesticación sucede también en el campo de la reflexión política donde surgen las normas de organización de lo que él llama “parques humanos”.¹⁷

¹⁵ Se trata del conjunto de técnicas que tienen como objetivo modificar y optimizar el comportamiento humano. Hay técnicas aplicadas por unos hombres sobre otros; y técnicas que los individuos aplican sobre sí mismos.

¹⁶ Este trabajo no tiene interés en profundizar en la crítica que hace Sloterdijk a Heidegger. Se trata de presentar los diferentes pensamientos del profundo debate acerca de la existencia humana en función de la técnica. Para Sloterdijk los hombres no tienen contacto con el mundo natural, es decir, el mundo natural no se nos da nunca en sí mismo, sino que disponemos técnicamente de él. Esto no ocurre solo desde la modernidad, sino desde los orígenes del hombre. No obstante, para Sloterdijk estar-en-el-mundo significa formar esferas, de tal manera que el-ser-en-esferas constituye la relación fundamental para el ser humano. En definitiva, se preocupa por los procesos a partir de los cuales el hombre se convierte en un ser que está en el mundo, y esta pregunta lo lleva a defender su tesis antropogenética.

¹⁷ Ibid, p. 75. En esta obra Sloterdijk tiene en cuenta el arte de la política en Grecia, entendido como el deseo de instruir a una minoría de hombres cuyo objetivo será el diseño de unas normas y conceptos racionales que les permitan gobernar a una mayoría analfabeta. Normas que son vistas por Sloterdijk como la gestión racional de parques humanos por parte de una élite domesticadora. Por eso, entiende la política como una técnica que no solo promueve el incremento de riquezas para el Estado (como afirmó Foucault), sino también favorece el mejoramiento del mundo. Sin embargo, mientras que las antropotécnicas más antiguas seleccionaron a unos pocos (chamanes, sacerdotes, magos, filósofos) para personificar el proyecto de ajustar las sociedades a las normas del mundo, utilizando como medio las letras y la religión; por otro lado, las antropotécnicas modernas reproducen en los individuos la capacidad para corregir los defectos del mundo a través del trabajo. En resumen, la antropotécnica como una mejora del mundo tiene que ver con las tecnologías del gobierno de las poblaciones, mientras que la antropotécnica de la mejora de uno mismo lo hace a través de las “tecnologías del yo” que habló Foucault. Según Sloterdijk ocurre un adiestramiento y crianza de unos hombres por otros; y también ocurre la crianza del individuo a partir de sí mismo. Ahora bien, lo

En relación con las anteriores observaciones sobre el humanismo, el término “posthumanista”² o *Humanidad Plus* surge, en un primer momento, del intento por superar la tradición humanista, aunque manteniendo algunos de sus elementos. La crítica de Sloterdijk al humanismo desde una perspectiva tecnológica implica replantear las concepciones humanistas anteriores. Según Sloterdijk hay que prescindir de una interpretación (humanista) que acepta la dicotomía sujeto-objeto. Además, el concepto de máquina necesita una revisión, ya que se origina en el contexto ontológico que distinguía las entidades con alma (las personas) y las entidades sin alma (las cosas, en este caso, también las máquinas). Así, la filosofía, nos dice Sloterdijk, no se ha preocupado demasiado por las entidades sin alma. Por este motivo, es imposible comprender la complejidad de la experiencia de nuestra época a partir de criterios lógicos que enfrenten, de manera excluyente, al hombre con la máquina.

En este debate es de utilidad recurrir a los planteamientos de Bruno Latour, quien se hizo célebre, junto a su teoría del actor-red, por cuestionar la asimetría metodológica y ontológica entre humano y no humano. Para Latour, lo que importa es la función que desempeña cada actor en una red, no el hecho de si es un ser humano, un ser vivo o un dispositivo tecnológico. Más en general, la teoría del actor-red también se opone al dualismo naturaleza-sociedad. Qué sea natural o social no es algo con lo que se cuenta desde el inicio para dar una explicación de un fenómeno, sino justamente el resultado de un proceso de creación y consolidación de una red. Cualquier entidad es el resultado de la coproducción de elementos naturales y sociales a diversos niveles de interacción. A la teoría del actor-red se le ha llamado también sociología de lo post-humano y post-social.¹⁸

que Sloterdijk quiere decir es que las antropotécnicas modernas ya no procuran prevenir al sujeto y protegerlo psíquica y corporalmente del sufrimiento de la existencia (como ocurrió durante mucho tiempo con las prácticas ascéticas); sino que buscan inducir a los individuos a vivir en una realidad donde predomina el riesgo.

¹⁸ Bruno Latour, *La esperanza de Pandora. Ensayos sobre la realidad de los estudios de la ciencia*. Ed., Gedisa, S.A, Barcelona (España), 2008.

El desarrollo tecnológico, sobre todo a partir de los años 80 del siglo pasado, permitió al hombre la posibilidad de trascender (en parte) su estado orgánico, con el fin de alargar la calidad de vida y mejorar enormemente la salud. A partir de este momento, surge el desarrollo de las nanotecnologías para la reconstrucción del cuerpo con la ayuda de nanorobots que, en el futuro, viajarán por todo el organismo a través de la sangre y repararán errores en el ADN de las células, o bien combatirán agentes patógenos. También aparecen nuevas investigaciones en criogenia para, por ejemplo, detener el envejecimiento celular. En suma, estos avances mencionados se fusionan con las tecnologías de la información, con el campo de la biotecnología y de las ciencias cognitivas. Es evidente, entonces, que estas tecnologías en desarrollo constituyen el camino hacia nuevos niveles de capacidad humana.

La utilización de la biotecnología, la ingeniería genética, los avances en IA y en robótica implican tal cambio en el mundo de la vida, que obligan a plantearnos el empleo de esta nueva posesión de conocimiento, de manera que el futuro de la especie humana no esté completamente en manos de la “transcendencia tecnológica”. El poder alcanzado actualmente por la tecnología es tal que se asemeja al poder que reservamos a los dioses. Los científicos predicen que el momento de crear IA que supere a la inteligencia humana es inminente. Por lo tanto, llegado el momento en el que se supere la inteligencia humana, será de gran ayuda tener en cuenta las repercusiones sociales que ocasionaría el descubrimiento de la replicación de la inteligencia humana en una IA. Es evidente entonces que hay que tomar medidas preventivas al respecto. Con todo, gracias a la manipulación tecnológica se alcanza un límite que nunca antes se había traspasado. Simplemente negar este nuevo paradigma resulta absurdo; también negar que la tecnología sea algo humano, ya que sería no aceptar la realidad humana, y con ello tampoco se resolvería el infinito debate de si la tecnología se sitúa en contra de la naturaleza humana o bien no existe tal naturaleza. Sirva de síntesis el argumento que a las críticas responde que todo lo que llega a hacer el hombre es también humano. La naturaleza humana no necesariamente ha de ser entendida como una esencia fija que permanece en el tiempo.

2. Estado actual

A continuación vamos a dar un salto desde la discusión de los filósofos anteriores a las ideas presentadas por autores (sean filósofos o no) que ofrecen argumentos a favor del desarrollo de la IA y de la robótica. Con la profetizada llegada de la superinteligencia se hará posible la promesa de suprimir la enfermedad, el dolor y el sufrimiento, e incluso la muerte. En definitiva, tendremos la esperanza de un “mundo feliz”. Sin embargo, cabe plantearnos si estamos o no preparados para afrontar los retos de un mundo en donde la superinteligencia conviva con el ser humano. Vemos cómo la realidad está consiguiendo derrumbar, al menos en parte, los límites de la ficción. En este contexto, algunos filósofos e incluso científicos del campo de la IA no están de acuerdo con la caracterización de la mente humana como una máquina. Otros en cambio, opinan que podemos ser capaces de entender nuestras mentes y cómo funcionan a fin de replicarlas mediante IA.

2.1. Una aproximación a la singularidad tecnológica

“¿Estamos preparados para ese cambio radical o bien pensamos que hay que conservar nuestro patrimonio genético y seguir siendo personas humanas, con nuestras limitaciones, pero conservando nuestra libertad y dignidad inalienables?”

CORTINA, Albert y SIERRA, Miguel Ángel, *¿Humanos o posthumanos? Singularidad tecnológica y mejoramiento humano*

Sobre el asunto se hace mención en el artículo titulado “retos de un futuro posthumano” de Albert Cortina y Miguel Ángel Sierra.¹⁹ Estos autores consideran la posibilidad de que el ser humano decida “voluntariamente” integrarse con las tecnologías NBIC (Nanotecnología, Biotecnología, Tecnologías de la Información y la Comunicación, y Neurociencia), con el objetivo de superar limitaciones tales como el envejecimiento, el sufrimiento y la enfermedad. Ambos defienden el transhumanismo desde un enfoque tecn-optimista.

Pese a que están de acuerdo en mejorar y perfeccionar las limitadas capacidades humanas, también consideran la importancia de analizar y reflexionar sobre los posibles riesgos y los retos que la futura superinteligencia causará en la sociedad. Por lo tanto, no solo tienen en cuenta los beneficios que suponen las mejoras transhumanistas de la “superinteligencia”, “superlongevidad” y del “superbienestar”, sino que además se plantean la cuestión de que tales mejoras sean favorables al bien común de la sociedad, y no en cambio, estén únicamente orientadas hacia los intereses particulares de unos pocos. No obstante, están de acuerdo con las principales tesis del transhumanismo y del posthumanismo, es decir, tienen en cuenta la idea de que cada uno de nosotros podría llegar a ser un humano mejorado o transhumano, y finalmente un posthumano. Según los autores esto podría ocurrir, ya que el deseo de perfeccionamiento es inherente a la naturaleza humana, si bien, --

¹⁹ Retos de un futuro posthumano: chips subcutáneos, métodos electromagnéticos para potenciar nuestro cerebro, prótesis externas. La incorporación de la tecnología a nuestro cuerpo y mente abre una nueva era. *El País*, España, 3 de enero de 2016. Albert Cortina, abogado y urbanista. Director del Estudio DTUM y Miquel-Àngel Serra, doctor en Biología. Gestor de investigación en la Universidad Pompeu Fabra. Coordinadores y autores del libro “*¿Humanos o posthumanos? Singularidad tecnológica y mejoramiento humano*” (Fragmenta Editorial, 2015).

apoyándose en la idea del biólogo Edward Wilson-- señalan que la esencia humana no reside tanto en la ciencia o en la tecnología, sino más bien en la cultura, en la educación y en las humanidades. En este sentido defienden una visión débil del posthumanismo, de manera que el ser humano aún tendría la responsabilidad de orientar y controlar la superinteligencia.²⁰

Por su parte, Ray Kurzweil, empresario y científico en ciencias de la computación e inteligencia artificial, en su obra *La singularidad está cerca*, afirma la hipótesis de que la IA alcanzará un nivel de inteligencia tal que no solo supere a la inteligencia humana, sino que llegado el momento, serán las máquinas inteligentes y no los seres humanos quienes se encarguen de su propia evolución y diseño, de modo que perderíamos el control de la situación. Por lo tanto, si aceptamos por completo la hipótesis de Kurzweil sobre la singularidad tecnológica, nos encontramos en la obligación de aceptar un fuerte determinismo tecnológico.²¹ Sin duda, la singularidad tecnológica implica también un cambio social, cultural, económico y político, en donde esta fusión tecnológica entre la cultura y la biología dará lugar a la replicación de la inteligencia humana en IA o en robótica y en humanos mejorados biológicamente o mediante dispositivos tecnológicos implantados (cyborgs).

Cabe preguntar si realmente podemos comprender el alcance de la creación de una superinteligencia. Para Nick Bostrom (director del Instituto para el Futuro de la Humanidad en la Universidad de Oxford) no es tan relevante predecir el momento de la singularidad, sino más bien cuestionar cuales serán los riesgos existenciales del futuro teniendo cuenta la singularidad tecnológica como un hecho innegable. Así afirma en una entrevista: “lo mejor que podemos hacer ahora mismo, en mi opinión, es impulsar y financiar el campo de investigación que se dedica a analizar el problema de controlar los riesgos futuros de la

²⁰ Cabe nombrar al catedrático de Lógica y Filosofía de la Ciencia en la Universidad de Málaga desde el 2010, Antonio Diéguez, autor de la reciente obra: *Transhumanismo. La búsqueda tecnológica del mejoramiento humano*. En esta obra se presentan las diferentes modalidades del transhumanismo tecnocientífico, señalando la importancia de la filosofía en cuanto a la reflexión ética y de valores sobre los riesgos y retos que presentan las nuevas tecnologías en desarrollo.

²¹ La teoría determinista propone que la tecnología fija el rumbo de las sociedades. En consecuencia, es frecuente encontrar representantes del determinismo tecnológico- como es caso de Jacques Ellul-, quien adopta una postura pesimista ante el avance de las nuevas tecnologías y el poder que conllevan. Tanto si el debate se sitúa completamente a favor de las tecnologías emergentes, como en contra, se observa un fuerte determinismo tecnológico. Es cierto que la tecnología influye enormemente en el proceso de la transformación de las sociedades; pero no será tampoco el único sentido de la existencia humana que haya que tener en cuenta a la hora de tener presente los retos y las asombrosas capacidades del nuevo momento tecnológico.

superinteligencia”²². Desde luego nos encontramos en un complejo debate entre quienes recurren a una especie de “humanismo renovado”, y quienes apuestan por un “humano plus”. Así, Fernando Broncano, catedrático de Filosofía de la Ciencia en la Universidad de Carlos III de Madrid, nos habla sobre dos visiones del posthumanismo: *el posthumanismo de la promesa* (la tecnología asume la potencia simbólica de la religión presuponiendo la obsolescencia de lo humano, siendo la superinteligencia el medio y el fin de la vida humana); y *el posthumanismo crítico* (que tiene en cuenta la capacidad reflexiva del ser humano para orientar y controlar el desarrollo tecnológico).²³

En relación con las afirmaciones de estos autores, uno de los problemas de la singularidad tecnológica es que no existe la posibilidad de pararse a reflexionar, ya que se asume la idea de que la tecnología mejora constantemente y que puede resolver cualquier problema de la realidad social. De este modo, las valoraciones y decisiones se postergan indefinidamente, o bien se dejan en manos de la propia evolución tecnológica. Por eso, nos encontramos ante un determinismo tecnológico cuyo fin es alcanzar una superinteligencia superior a la humana y no tanto un planteamiento crítico sobre el poder de la inteligencia artificial en la sociedad y como controlar sus riesgos.

Si bien, con el propósito de que el lector pueda comprender las aplicaciones en la sociedad de los recientes descubrimientos en IA y robótica, en este apartado también se tendrá en cuenta la convergencia tecnológica de la robótica con otros campos, como es el caso de la robótica cognitiva. En suma, este trabajo tiene el objetivo de vislumbrar la labor de la filosofía con respecto a la reflexión y aceptación de la IA y la robótica como un importante elemento en la sociedad del futuro.

²² El reto de las máquinas pensantes. *El MUNDO*, Madrid, España, 18 de enero de 2015.

²³ Conferencia de Fernando Broncano “Dos formas de posthumanismo: obsolescencia de lo humano u obsolescencia de las categorías”, realizada en la Universidad de la Laguna en el año 2015.

2.2. La convergencia tecnológica en la robótica

La robótica y la IA han alcanzado un punto decisivo en su evolución. Un robot, lejos de ser considerado un artilugio mecánico, logrará en un futuro, si todo sigue según lo previsto, crear su propio linaje y controlar su evolución. Los nuevos robots inteligentes autónomos serán capaces de razonar, percibir, actuar ante entornos dinámicos e imprevisibles por sí mismos. En la actualidad hay diferentes campos de investigación sobre robótica avanzada: robótica basada en el comportamiento o la conducta; robótica cognitiva; robótica de desarrollo o epigenética; robótica evolutiva; y robótica biomimética. La robótica inteligente autónoma consta de sistemas dinámicos que consisten en un controlador electrónico acoplado a un cuerpo mecánico. Estos robots necesitan de un sistema sensorial adecuado para percibir el entorno donde se desenvuelven. Además cabe aclarar que la robótica abarca un gran campo de estudio interdisciplinar entre la ingeniería mecánica, la electrónica, la informática, la física, anatomía, biología, zoología, etología y psicología, entre otras disciplinas. Las actuales investigaciones en robótica tienen como fundamento la ciencia cognitiva corporizada y la nueva inteligencia artificial. En este apartado se tendrá en cuenta únicamente la robótica cognitiva. Estos robots disponen de una estructura mecánica adaptable con el fin de garantizar una cierta destreza física-locomotora; de complejos sistemas efectores para llevar a cabo las tareas asignadas; y de sistemas de control para poder efectuar acciones correctivas. Por esta razón, los estudios sobre la cognición humana también son aplicados a la robótica utilizada en entornos muy complejos, cambiantes y dinámicos. Para que estos robots puedan actuar en entornos imprevisibles, deberán realizar funciones de cognición tales como razonar sobre las metas, las acciones, el tiempo, los estados cognitivos de otros robots, de manera que puedan aprender de la experiencia. A su vez, este aprendizaje a partir de la experiencia implica la interacción con humanos.²⁴

Francisco Varela fue un importante biólogo que desarrolló numerosos trabajos en el campo de la neurobiología, la biología teórica, la epistemología y especialmente la neurociencia cognitiva²⁵. El estudio de la cognición humana realizado por Varela está

²⁴ MEGATENDENCIAS, *Los robots inteligentes autónomos son la nueva generación. Están situados en su entorno, adoptan comportamientos, razonan, evolucionan y actúan como seres vivos*, España, 5 de octubre de 2005.

²⁵ Fue director de investigación en el centro Nacional de Investigaciones científicas de Francia y director del laboratorio de Neurociencias Cognitivas en el Hospital de la Pitié-Salpêtrière en París.

vinculado con la cibernética. En este sentido, la cibernética actual es definida como la imitación a través de máquinas de los sistemas de regulación automática de los seres vivos. Sin embargo, Varela se sitúa en contra de los pioneros de la cibernética que consideraban a la lógica como la disciplina adecuada para comprender el cerebro y la actividad mental. Sobre esta base lógica se construyeron las primeras computadoras. Este autor nos dice al respecto: “la intuición central era que la inteligencia (incluida la inteligencia humana) se parece tanto a un ordenador en sus características esenciales, que la cognición- el conocimiento- se puede definir como la computación de representaciones simbólicas”.²⁶ Por el contrario, para Varela la cognición es un proceso “enactivo”. La designación de “enactivo” proviene del verbo “to enact” que significa “poner en ejecución”, “representar” o “actuar”. Así afirma Varela: “la cognición no es entonces la representación de un mundo pre-dado por una mente pre-dada, sino más bien la puesta en obra de un mundo y una mente a partir de una historia de la variedad de acciones que un ser realiza en el mundo”²⁷. Con esto quiere decir, que el conocimiento es acción en el mundo (ejecución) y no representación del mundo. El conocimiento se refiere a una situacionalidad.

Es el caso de la robótica cognitiva, que utiliza técnicas procedentes del campo de las ciencias cognitivas. Resulta evidente que el área de aplicación de los robots cognitivos es la interacción con humanos, se trata de robots que pretenden actuar como humanos. Un ejemplo de interacción sería el habla, por lo tanto estos robots también están dotados de un tipo de modelo simbólico. Sirvan de ejemplo las investigaciones llevadas a cabo en el Centro de investigación Conjunta de la Unión Europea- ubicado en Ispira, Italia- del científico informático español José Millán. Durante años se ha encargado de enseñar a una pareja de robots a aprender por sí mismos, con el fin de que en un futuro próximo, encuentren un puesto de trabajo en la vigilancia o el manejo de material en las centrales nucleares.

Lo anteriormente expuesto no solo sirve para comprender el fundamento de los nuevos avances en robótica, sino también para observar la tendencia por parte del humano a imitar seres vivos (incluido el ser humano) a través de artilugios. A este deseo se le ha denominado recientemente inteligencia artificial (IA).

²⁶ REVISTA CHILENA DE NEURO-PSIQUIATRÍA.v.39 n.4 Santiago oct. 2001.

²⁷ REVISTA CHILENA DE NEURO-PSIQUIATRÍA.v.39 n.4 Santiago oct. 2001.

3. Discusión y posicionamiento

Seguidamente analizaremos los retos que la robótica plantea desde un punto de vista ético-político, es decir, en quién recae la responsabilidad de las decisiones tomadas por los robots, y cómo lograr que estas decisiones sean las más apropiadas desde un punto de vista de los valores socialmente aceptados. Otra tarea consistirá en decidir cómo regular sus posibilidades de autonomía y de interacción con los seres humanos. Para abordar tales cuestiones se tendrá en cuenta la propuesta del parlamento europeo de que los robots paguen impuestos, como medida preventiva para paliar la enorme tasa de desempleo provocada por la “Segunda Era de la máquina”.

3.1. Responsabilidad

“La Segunda Era de la máquina” --llamada de esta manera por el profesor Erik Brynjolfsson del MIT (Instituto Tecnológico de Massachusetts)-- implica un nuevo paradigma de enorme complejidad y confusión. En el parlamento europeo, Mady Delvaux (Vicepresidenta de la Comisión de Asuntos Jurídicos y miembro de la Delegación en las Comisiones Parlamentarias de Cooperación de la UE) ha presentado un informe, aprobado por Bruselas en enero de este año, sobre las reflexiones y sugerencias suscitadas por el nuevo estatus legal otorgado a los robots.²⁸ Plantea la necesidad de una legislación con relación a la responsabilidad y la seguridad, aplicable tanto a las empresas que los fabrican, como a las que utilizarán este tipo de tecnología. Teniendo en cuenta que en la actualidad Europa lidera el mercado de la robótica, y para seguir manteniendo este estatus, según Delvaux los futuros consumidores contarán con medidas cautelares sobre la seguridad y la protección de datos. Este tipo de medidas serán aplicables sobre todo a los robots de asistencia o de entretenimiento, coches autónomos, robots industriales (se matiza que “no estamos hablando de armas”). Actualmente los robots autónomos son capaces de subir información a la red con el fin de instruir y enseñar a otros robots. Así pues, habrá que controlar quien tendrá acceso a tales datos y si se formulará una ley de protección de datos en robótica. Será preciso en este trabajo reparar en la autonomía e interacción de los robots en la cotidianidad humana, si bien

²⁸ Delvaux propone normas europeas para la robótica y un seguro obligatorio para los modelos grandes, *Noticias Parlamento Europeo*, Bruselas, 12 de enero de 2017.

estas cuestiones se atenderán en el punto de este apartado que trata sobre la interacción y la autonomía en robótica.

Volviendo la mirada hacia la responsabilidad y seguridad de la robótica, en caso de accidente se contemplan dos opciones: la primera, el fabricante sería el responsable directo, aunque también se podrá pedir cierta responsabilidad a sus proveedores; y la segunda, sería una evaluación de los riesgos teniendo en cuenta un sistema de compensación entre las partes implicadas, e incluso la obligatoriedad de un seguro, “al menos para los robots grandes” afirma Delvaux. Si bien es cierto que los avances en robótica aplicados a la industria, el mercado y el empleo ayudarán a crear puestos de trabajos en ciertos sectores, también puede destruirlos en otros. Al respecto, Delvaux también admite que los robots no sustituirán por completo a los humanos en el trabajo, sino que habrá una cooperación entre ambos. Según esta política, cabe preguntarse cómo afrontar esta problemática y ver qué tipo de tareas llevarán a cabo los robots; por eso considera la importancia de analizar estos aspectos con el fin de que llegado el momento en el que los robots autodidactas sean una realidad estemos preparados para afrontar los posibles daños o riesgos causados por la interacción entre este tipo de robots y los seres humanos en la vida social. A modo de solución propone dar a los robots un tipo de “personalidad electrónica” limitada. Por este motivo, pretende crear un marco legal que regule las decisiones y las acciones realizadas por los robots. En resumen este marco legal, además deberá incluir qué tipos de trabajos serán destinados a los robots y el papel que desempeñarán en la sociedad. Delvaux considera que estas cuestiones suscitan una profunda reflexión sobre cómo queremos vivir y trabajar y cómo interactuar con los robots.

Llegados a este punto, resulta evidente la controversia causada por estas declaraciones. Así, nos encontramos con expertos que no apoyan la implantación de esta medida, ya que si se aplica a la situación social actual, desencadenaría una mayor desigualdad y precariedad en el trabajo. Si la solución es dotar a los robots de cierta capacidad jurídica, ello implica que las empresas se encargarán de dicho impuesto. Se puede suponer entonces que sea más rentable para los empresarios contratar a un robot que a un ser humano, ya que por ejemplo un robot no necesitaría un sueldo, pagas extras, una jornada laboral, vacaciones, etc. Por lo tanto, en un principio me parece que esta medida favorece mucho más a las empresas que a los trabajadores humanos. Según un estudio del Foro Económico Mundial, se estima la pérdida de 7 millones de empleos en cinco años. Además, otras investigaciones

sitúan una tasa de desempleo de un 60% en la manufactura, la agricultura y los servicios. Esto también supone una desigualdad en el sistema económico mundial basado en el consumo, ya que si tenemos en cuenta los datos anteriores sobre el desempleo, el consumo disminuirá de manera considerable. Nos encontramos ante un cambio de paradigma que implicará importantes modificaciones con respecto al trabajo y la vida del hombre. La cuestión será si estamos o no preparados para afrontar un cambio tan significativo en la dimensión social del ser humano. Cabe también preguntarse dónde se invertirá el dinero de los impuestos de los robots. Ese dinero podrá ser destinado según los expertos a los servicios públicos o al sistema de pensiones; o bien será utilizado a modo de renta básica universal, como también se ha propuesto. Actualmente vemos que el dinero se ha transformado virtualmente. El bitcoin es una moneda virtual que facilita un nuevo sistema de pago a través de una red descentralizada, es decir, no hay ninguna autoridad central o ningún intermediario que lo respalde. En otras palabras, se trata de un tipo de dinero utilizado por los usuarios de internet. En relación a su funcionamiento, la red bitcoin cuenta con una contabilidad pública llamada “block chain”, de manera que esta contabilidad abarca todas las transacciones procesadas y permite, a su vez, la verificación de la validez de cada una de las transacciones. Las transacciones quedan protegidas por un sistema de firmas digitales correspondientes a las direcciones de envío. En efecto, la robótica, la IA, y la biotecnología podrían en un futuro valerse de este nuevo sistema de pago.

En la actualidad existen empresas como *Bell industries* que se dedica a la fabricación de las piezas que posteriormente son utilizadas para la construcción de robots con forma humana; es el caso de *Bostom Dynamics* que fabrica robots antropomorfos para mano de obra; o bien *Bae Systems*, que se encarga de fabricar sistemas de inteligencia y robots aplicados a la armamentística. Por lo tanto, urge una reflexión interdisciplinar y crítica sobre estas cuestiones, de forma que podamos ser capaces de orientar el desarrollo de estas tecnologías, teniendo en cuenta tanto los beneficios como los riesgos que supone la interacción de la robótica avanzada en la sociedad. Por esta razón, será necesaria una normativa con marco legal que regule el uso de la robótica por parte de las empresas, la cuestión será ver si una normativa o ley será capaz de afrontar los cambios sociales que se avecinan. Sin ir más lejos, la empresa *Bell industries* utiliza los últimos avances científicos y tecnológicos sobre IA y robótica con el fin de crear armas destructivas. ¿Por qué se permite esto? Se supone entonces la idea de que el beneficio de la práctica científica no se realice en función del poder económico y político; sin embargo, las empresas y la política controlan la

toma de decisión científica y tecnológica en cuanto a las investigaciones y el uso de esta tecnología en la sociedad.

3.2. Autonomía de los robots e interacción con los seres humanos

Son conocidas las tres leyes de la robótica mencionadas por Isaac Asimov²⁹, sin embargo con la llegada de la IA y los nuevos avances en robótica autónoma, y en consecuencia la nueva oleada de automatización de la vida privada y social del ser humano, se ha propuesto una cuarta ley. Hans Moravec es un investigador en IA y robótica, en general, es conocido por sus estudios sobre los impactos de la tecnología en la sociedad. Propone entonces la siguiente ley: “ un robot debe reproducirse, siempre que dicha reproducción no interfiera con la primera, la segunda o la tercera ley”. Esta última ley pretende hacer frente a un futuro en el que los robots podrían construir su propio linaje y guiar su evolución. La IA y la robótica permite la posibilidad de crear coches autónomos; aviones militares pilotados por una IA; drones y robots antropomórficos para fines bélicos. Llegado este momento, hay que plantear cuestiones biopolíticas para considerar- tal y como se ha propuesto en el parlamento europeo- en quien recaerá el asunto de la responsabilidad como de la autonomía para la toma de decisiones por parte de estas inteligencias.

Hoy día se está estudiando la posibilidad de programar robots para que sean capaces de tomar ciertas decisiones morales. Por ejemplo, los nuevos coches autónomos de la empresa Tesla Motors o los nuevos coches autónomos de Google, plantean profundos debates éticos. Imaginemos el siguiente supuesto: Philipa Foot fue una filósofa británica que propuso el dilema ético del tranvía, si aplicamos este dilema a un posible escenario en donde estén implicados un coche autónomo en el que viaja una persona, y resulta que dadas las circunstancias, lo único que puede hacer el coche autónomo para evitar atropellar a otras cinco personas que se encuentran en medio de la carretera, es girar a la izquierda y caer por un precipicio, ¿qué debería decidir hacer el coche autónomo?, ¿debería aplicar un pensamiento utilitarista y decidir salvar a las cinco personas en lugar de su ocupante? Por

²⁹Un robot no debe dañar a un ser humano o, por su inacción, dejar que un ser humano sufra daño; 2)Un robot debe obedecer las órdenes que le son dadas por un ser humano, excepto cuando estas órdenes se oponen a la primera Ley; 3)Un robot debe proteger su propia existencia, hasta donde esta protección no entre en conflicto con la primera o segunda Leyes. *Manual de robótica*, 56.ª edición, año 2058.

otro lado, la perspectiva deontologista sostiene que hay normas o principios morales que hay que cumplir incondicionalmente, pero si aceptamos esta corriente ética, el coche autónomo cumplirá por ejemplo con la máxima “matar está siempre mal”, y entonces acabaría atropellando a las cinco personas. Al respecto, lo importante será cuestionar quién será el encargado de elegir el criterio ético: ¿los programadores?, ¿el mismo usuario?, ¿la empresa o bien el estado? Actualmente estas cuestiones se encuentran en pleno debate e incluso, los expertos se encuentran con el problema de que los robots tardan en tomar decisiones éticas, debido a que pasan demasiado tiempo en decidir. Esto plantea la cuestión de si se podrá programar la intuición a la IA y los robots para que estos puedan tomar una decisión comprometida. En este trabajo no interesa vislumbrar cual será la solución para solventar los dilemas éticos que se presentan, sino más bien hacer una aproximación de los mismos para comprender que mientras más sofisticada sea la IA y la robótica, mayor será la responsabilidad y la autonomía que adopten estas tecnologías, y por lo tanto, los límites de su programación para la interacción con seres humanos, serán aun más complejos. La solución no solo estará en manos de unas leyes o normas, o de un botón que en caso de peligro desconecte a una IA o a un robot fuera de control. Así se pretende conseguir que las máquinas no tengan autonomía plena sino que haya un control humano, ¿pero qué humanos tendrán ese control? En Estados Unidos, la organización DARPA (La Agencia de Proyectos de Investigación Avanzados de Defensa), está desarrollando el programa CoABS, se trata de un sistema de gestión y coordinación de tropas diseñado para reducir el tiempo en la toma de decisiones. Este sistema trabaja en red de manera que si la toma de decisiones está distribuida a lo largo y ancho de una red, seguirán funcionando aunque algunas de sus partes fueran destruidas. Esta idea dio lugar a internet, por este motivo, si ya se está experimentando con robots que son capaces de conectarse a una red de internet, queda claro que un botón u otro dispositivo de control se vuelve mucho más complejo de formular. Uno de los mayores riesgos, se encuentran en que las grandes potencias militares ya llevan tiempo desarrollando y aplicando los nuevos avances en IA y en robótica. Recientemente la industria armamentística soviética ha anunciado que su cazabombardero (sukhoi T-50), tendrá una versión pilotada por una IA. Cabe aquí decir que estas nuevas tecnologías armamentísticas implicarán un mayor negocio, ya que los soldados humanos si sienten empatía hacia otras vidas, se sienten culpables al matar, tienen incluso estrés postraumático, sus bajas en el campo de batalla implican enormes pérdidas económicas, pero en cambio un robot humanoide o un dron, podrán cumplir los objetivos bélicos sin sentir remordimiento alguno por sus acciones. Como respuesta, un grupo de intelectuales, entre los que se encuentra Noam Chomsky o Jody

Williams (activista y Premio Nobel de la paz), han puesto en marcha una campaña en contra de la fabricación de robots que atenten contra la vida humana. Con todo, resulta importante preguntarnos hacia dónde se dirige el poder de la ciencia y de la tecnología, cuales son sus medios y fines.

3.4. Tecnología y Civilización

Ritos, arte, poesía, drama, música, danza, filosofía, ciencia, mitos, religión, son todos los componentes esenciales del alimento cotidiano del hombre, pues la auténtica vida de los seres humanos no solo consisten en las actividades laboriosas que directamente lo sustentan, sino también en las actividades simbólicas que dan sentido tanto a los procesos de su quehacer como a sus últimos productos y consecuencias

LEWIS MUNFORD, *El mito de la máquina*
(*Técnica y evolución humana*)

Para defender la tesis mencionada en este trabajo, propongo una reflexión general sobre hasta qué punto el desarrollo científico supone una mejora de los problemas sociales, o si bien, a menudo, significa todo lo contrario. En este apartado, se sugiere que se tengan en cuenta las principales reflexiones sobre las consecuencias devastadoras de la tecnología. Asimismo, hay que mencionar la ambivalencia de la tecnología entre las tendencias civilizatorias y bárbaras de las sociedades actuales; todo ello en contraposición con la concepción de la tecnología como una aplicación del conocimiento científico, en donde el progreso vendría dado por un aumento del grado de civilización en la sociedad.

El ideal del positivismo lógico era que solo existe una ciencia y un método independientemente del objeto al que se pudiera aplicar; pues bien, frente a esta visión, en el siglo XIX sucede una reacción que defiende la autonomía de las ciencias sociales frente a las ciencias naturales. En las ciencias sociales intervienen motivos, valores, deseos, que tenemos que tener en cuenta y comprender para luego explicar. Esto determina la autonomía de las ciencias sociales frente a las naturales. Por otro lado, la ciencia de principios del siglo XX defendía la idea de que los factores externos no repercutían en el desarrollo científico. Así, los problemas de la ciencia y la finalidad de su conocimiento sólo eran entendidos bajo una perspectiva teórica dentro del contexto de justificación de las teorías. Sin embargo, cuando un científico determina el éxito o el fracaso de una teoría lo hace desde un entramado político, económico y social. Hay numerosos estudios realizados desde la filosofía que tienen en cuenta la relación entre ciencia, tecnología y política (incluyendo sus relaciones con el estado). Por lo tanto, se defiende la convicción de que la teoría se concibe desde un mundo de vida, o dicho en otras palabras, desde la praxis.

En 1924 se funda el Instituto de Investigación Social, pero no será hasta 1931 cuando Horkheimer asuma su dirección, momento que coincide también con la creación de la revista de Investigación Social. Después de la primera guerra mundial se reúnen un grupo de intelectuales para debatir sobre la historia, conciencia de clases, marxismo y filosofía. Elaboraron una teoría crítica para que no asumiese únicamente las tesis marxistas, sino que las interpretase desde el presente con el fin de transformarlo. En realidad, la *Teoría Crítica* es la integración y reinterpretación de varios enfoques (Marx Weber, Marx, Nietzsche, Freud, etc), con el fin de transformar la realidad. Además, la teoría crítica de la escuela de Frankfurt está considerada por sus integrantes como una vertiente crítica de la tradición ilustrada.

Tanto en el inicio de la Teoría Crítica como en los comienzos de la Dialéctica de la Ilustración existe una experiencia histórica dolorosa. En la Dialéctica de la Ilustración, Adorno y Horkheimer reparan sobre este asunto. Así, en el prólogo sostienen: “nos habíamos propuesto era nada menos que comprender por qué la humanidad, en lugar de entrar en un estado verdaderamente humano, se hunde en un nuevo género de barbarie”.³⁰ Estos autores concluyen que en el seno del pensamiento ilustrado está presente una enorme contradicción. Por este motivo, el título de la obra utiliza el concepto de dialéctica negativa, ya no entendido desde el sentido griego que remitía al diálogo, sino desde la contradicción, creándose una especie de tesis y de antítesis al mismo tiempo. Es preciso traer a debate que el contexto de esta época estuvo marcado por el auge y la expansión del nazismo, perpetuándose un pensamiento totalitario que termina por recaer en la barbarie. Por este motivo, Adorno y Horkheimer se plantean explicar cómo el modelo ilustrado que prometía las mejores metas para la humanidad acaba desmoronándose.

Para estos autores la única forma de abordar el problema del conocimiento es poniéndolo en relación con la sociedad; la producción del conocimiento es, para ellos, una actividad más de la sociedad, una práctica social. La teoría no es algo aislado, se inserta dentro de un contexto social e histórico determinado que va a influir en la manera en la que se produce el conocimiento. Evidentemente el sujeto de conocimiento dejaría de ser un sujeto abstracto (separado incluso del cuerpo, como en la filosofía cartesiana); y pasa a ser un sujeto concreto, empírico, determinado por esa situación histórico social en la que vive. Se trata, por

³⁰ HORKHEIMER, M., ADORNO, T., *Dialéctica de la Ilustración. Fragmentos filosóficos*. Ed. Trotta, 2009.

tanto, de un ser humano concreto que está en una constelación de relaciones sociales que van a influir y a determinar la forma, el modo, incluso el contenido de la actividad teórica.

En la Dialéctica de la Ilustración, la crítica que se hace del concepto de Ilustración tiene como objetivo preparar una lectura positiva de la misma. Esto es planteado desde una doble tesis: el mito es ya ilustración; y la ilustración recae en mitología. Con ello, se sitúan en contra de la idea clásica de progreso que avanza desde el mito al logos. Su propuesta es que el mito es también logos, es decir, la mitología es también una forma de ordenar el mundo en una disposición con un sentido, una finalidad. Hay que tener en cuenta que la Ilustración nace bajo el signo del dominio, siguiendo un programa de desencantamiento del mundo. A partir de entonces, el fin del conocimiento se desvincula de la búsqueda de la felicidad, y aspira a la explotación y al dominio de esta naturaleza desencantada. Ya desde Bacon se abandona el saber como deleite y se plantea el conocimiento como una forma de poder. Descartes, por su parte, en el *Discurso del Método*, también tiene en cuenta la importancia de elaborar un método universal con el que alcanzar el dominio de la naturaleza. Adorno y Horkheimer defienden la idea de que esta tendencia al dominio (entendido este como una patología de la razón), está ya presente en el mito. Así al convertirse la ilustración en un mito ésta queda reducida a razón instrumental. Lo que viene a decir la teoría crítica es que la actividad teórica forma parte de la praxis, está envuelta en un contexto mayor, en una praxis general y, por tanto, no puede situarse exclusivamente en el plano de la mera contemplación.

Retomando de nuevo la reflexión sobre el concepto de civilización del siglo XX hay que dejar claro la idea de que la nueva ciencia que se venía desarrollando constituía la mejor expresión del pensamiento racional. Jacques Ellul, en su obra *la Técnica o la apuesta del siglo*, nos dice que la utopía (el fin de la enfermedad, de la guerra, de la escasez) acaba siendo una idea respetada por la ciencia. En este contexto se enmarca la idea de progreso como la utopía de la humanidad, que considera que cualquier avance tecno-científico implicará por sí mismo una mejora en las condiciones materiales y sociales de vida. El poder de la técnica puede ser entendido como una consecuencia de las relaciones de poder entre la ciencia y el estado, esto recibe el nombre de tecnología social. Así, el poder no depende únicamente de los artilugios tecnológicos, sino también de procedimientos organizativos.

Lewis Mumford, quizás el pensador por excelencia de la técnica en el siglo XX, se sitúa dentro de la concepción humanista, otorgando prioridad al ser humano para alcanzar una

interpretación del significado de la tecnología, de tal manera que se produzca una interdisciplinariedad y armonía entre la empresa tecnológica y el saber humanista. El pensamiento de Mumford no aboga por un rechazo simplista de la tecnología, sino que establece una distinción entre dos tipos de tecnologías: las democráticas (que están de acuerdo con la naturaleza humana), y las autoritarias (que mantienen un conflicto, muchas veces violento, en contra de los valores humanos). En la propuesta de Mumford es fundamental el concepto de “Megamáquina”³¹. Alude a un sistema técnico que promueve la despersonalización y el absolutismo, ya que su desarrollo se basa solamente en la acumulación de poder. Así, las características de la “Megamáquina” son el orden, la obediencia, la ejecución automática, la desigualdad, la esclavitud y la eficiencia. Todas ellas favorecen la ignorancia respecto a las necesidades y fines de la vida, y por tanto, se basan en una tecnología autoritaria donde predominan los aspectos negativos, tiránicos y destructivos, dado que su objetivo es aumentar la eficacia de la “Megamáquina” mediante métodos totalitarios que impiden la función activa y libre de la persona humana. Para Mumford lo relevante está en cambiar este tipo de tecnología autoritaria por una tecnología democrática que permita la participación activa del ser humano en todos sus ámbitos.

Mumford plantea un paralelismo entre lo orgánico y lo técnico, superando los anteriores enfoques mecanicistas. En el capítulo llamado *El nuevo Órganon*³², desarrolla una interesante propuesta: que la relación entre la tecnología y la inventiva humana no se debe resolver con la exaltación de los objetos y artefactos. Adorno y Horkheimer denominan esta exaltación como la razón instrumental. A diferencia de estos autores, según Mumford la solución tampoco estará en una razón ampliada o matizada. Para Mumford la única forma de contrarrestar los horrores del “pentágono del poder”³³ es centrarse en los aspectos positivos de la biotécnica y en la creatividad, que favorecen la capacidad de la sociabilidad. De lo contrario, la ciencia, las instituciones sociales y la política contribuirán a la expansión del poder potenciando un sistema aún más deshumanizado.

³¹ Véase Mumford, Lewis, *El diseño de la megamáquina*. En su: *El mito de la máquina. Técnica y evolución humana*. Ed. Pepitas de calabaza, España, 2010, pp. 311-341.

³² Véase Mumford, Lewis, *El nuevo Órganon*. En su: *El pentágono del poder. El mito de la máquina (dos)*. Ed. Pepitas de calabaza, España, 2011, pp. 613-637

³³ *Ibid.*, p.268

En la obra *El mito de la máquina. Técnica y evolución humana*, Mumford nos habla sobre las transformaciones a gran escala en el entorno humano debido a los avances de la matemática y la física aplicados a la tecnología. El descubrimiento de la energía nuclear, la cibernética y las teorías de la información han impulsado una transformación de la conducta humana, de manera que si no somos capaces de comprender los riesgos y los impactos que suponen, dicha transformación podrá ser aún más extrema. En este sentido, el hombre no sólo ha alcanzado un momento de su historia que le ha permitido controlar y dominar la naturaleza, sino que además comienza a distanciarse de su mundo orgánico. Mumford se plantea las consecuencias y los riesgos que tendrá este distanciamiento. La nueva técnica es denominada por Mumford como “megatécnica”. Al respecto, opina que esta “megatécnica” convertirá al hombre en un animal pasivo que actuará según los fines de la máquina, de tal manera que su individualidad y su subjetividad quede totalmente olvidada y silenciada. Las implicaciones del pensamiento de Mumford son de importancia en relación con nuestro compromiso con las actuales formas de progreso científico y técnico. Desde que Mumford escribía sobre estos temas, vemos cómo ha aumentado no sólo el control tecnológico sobre el entorno natural, sino también en la dimensión social. Según Mumford hay motivos suficientes para pensar que la mejora de la sociedad no ocurrirá cuando se alcance la supremacía de las máquinas. Cree que ha habido y que persiste un exceso de dependencia tecnológica, que a su vez se debe a una interpretación errónea de la técnica. Habría una confianza ciega en la ciencia y la tecnología para salvar a la humanidad. Esto quiere decir que hemos estado equivocados al considerar la capacidad técnica como un referente de inteligencia, ya que la fabricación de herramientas no fue algo singularmente humano hasta que no estuvo modificada por símbolos lingüísticos y artísticos como conocimientos socialmente transmitidos. Entonces lo que marcó la diferencia no fue la mano del hombre, sino su cerebro y mente. Al respecto nos dice Mumford: “mediante la tenaz exploración que el hombre hizo de sus capacidades orgánicas, se asignaron nuevos papeles a ojos, oídos, nariz, lengua, labios y órganos sexuales. Hasta la mano dejó de ser, como antes, una herramienta especializada: ahora acariciaba el cuerpo amado, estrechaba el bebé contra el pecho, hacía gestos significativos o expresaba en rituales compartidos y danzas preestablecidas sentimientos de otro modo inexpresables acerca de la vida o de la muerte, de un pasado documentado en la memoria o de un futuro preocupante. Por tanto, la técnica de las herramientas no es más que un fragmento de la biotécnica, de la dotación vital del

hombre.”³⁴ El propósito de estas páginas también consiste en reflejar que el progreso de las grandes civilizaciones, también incluye la barbarie. Prueba de ello es el proyecto Manhattan, programa nuclear encubierto de EEUU en el que participaron 13 mil empleados, cuyo objetivo era superar a los científicos alemanes que formaban parte del Proyecto Uranio dirigido por Ernest Einsenberg. En el proyecto Manhattan también estaban involucrados científicos como Enrico Fermi (construyó el primer reactor del mundo); Robert Oppenheimer (fue el director del proyecto); Jonh Von Neuman (era el experto en materia de explosivos); Seth Neddermeyer (fue el creador de la primera bomba nuclear de la historia, Trinity). La bomba que fue lanzada en Hiroshima el 6 de agosto de 1945 era el mismo modelo de bomba que había propuesto Neddenmeyer, y terminó matando a 140 mil personas. El 9 de agosto se lanzó otra bomba en Nagasaki. Esta bomba mató a 40 mil personas, pero se cree que falló en su detonación porque tenía una potencia mucho mayor. Este proyecto además reunió a importantes científicos, entre ellos Albert Einstein y Richard Feynman. La cuestión más importante sería comprender por qué la ciencia termina por aplicarse a los fines más macabros, y cómo los científicos más “brillantes” participaron en este proyecto de destrucción. Es un trágico ejemplo de una ciencia y tecnologías al servicio del Estado y de lo militar. Hay otras expresiones de la barbarie como son los medios de exterminio legal, las técnicas modernas de tortura física y psicológica, la experimentación con seres humanos o las medidas eugenésicas. Por otra parte, nos encontramos con otras formas más sutiles de barbarie que utilizan técnicas para someter a los individuos y castigarlos si no cumplen con los modelos determinados de comportamiento (en la educación, en las prisiones, en el trabajo). En cualquier caso, no es de extrañar que la sociedad actual reciba el nombre de sociedad del riesgo. No queda claro que seamos la especie más inteligente, pero sí la más peligrosa. ¿Cómo es posible que la vida en su complejidad se haya desarrollado hace miles de millones de años, y nosotros los humanos, podamos acabar con su equilibrio en menos de 200 años?

Cuando seamos capaces de crear una IA que supere a la inteligencia humana, el destino de nuestra especie estará en manos de la superinteligencia. Nick Bostrom, en su obra *superinteligencia: caminos, peligros y estrategias*, reflexiona sobre cómo afrontar un futuro en el que la inteligencia artificial supere a la humana. Bostrom se pregunta si la humanidad será capaz de controlar y orientar el desarrollo tecnológico de la singularidad tecnológica.

³⁴ Véase Mumford, Lewis, El mito de la máquina. Técnica y evolución humana. Ed., Pepitas de calabaza, España, 2010, p. 16

Además considera la idea de que la superinteligencia pueda incluso acabar con la vida humana en el planeta. Sin embargo, nos encontramos en un momento en el que la inteligencia artificial no ha superado a la humana, y en consecuencia, podemos prepararnos para afrontar tal singularidad. En este momento el ser humano todavía puede anticiparse a sus propias creaciones. Bostrom aborda este asunto partiendo desde el estado actual del mundo, la ciencia y la tecnología, dentro de un enfoque filosófico, ético y humano. El objetivo de su obra es afrontar la complejidad del desarrollo tecnológico y cada detalle de esta complejidad. Para reducir los riesgos de la IA propone una visión estratégica, es decir, una especie de plan de acción. También propone como objetivo específico promover las buenas prácticas entre los investigadores de la IA. Bostrom considera que nos encontramos a punto de entrar en un cambio que puede compararse con la aparición de la vida en la tierra, y es que estas tecnologías evolucionan cada vez más rápido, de manera que apuntan hacia un objetivo común, la superación de los límites humanos. Por lo tanto, se produce una aceleración de la información, de manera tal que la información se duplica el doble de rápido en cada generación. En definitiva, estamos en la cúspide de un salto evolutivo radical hacia un horizonte de complejidad. Por este motivo, Bostrom nos habla sobre los caminos y las estrategias a seguir, para enfrentarse a los riesgos y a los beneficios que conllevan la creación de una superinteligencia.

La IA puede ser definida, en un primer momento, como el intento por parte de filósofos y científicos de responder al enigma de cómo el cerebro humano da origen a pensamientos, emociones y conciencia. La IA tendría su origen en la suposición de que el cerebro humano sería como una máquina. Hay una escuela de pensamiento que considera la mente-como-máquina, ya que el pensamiento se lleva a cabo en el cerebro humano. Por otro lado, la escuela opuesta, la-mente-más-allá-de-la-máquina, critica el enfoque que aplica únicamente la lógica a la capacidad de pensar. En este trabajo no voy a entrar en el debate sobre la naturaleza de la conciencia. Sin embargo, considero que existe una importante línea de investigación que conecta la IA y la robótica, por un lado, y el campo de la filosofía de la mente, por el otro. De acuerdo con esta última hipótesis, la definición más citada de la IA es la establecida por Raymond Kurzweil: “la inteligencia artificial es el arte de crear máquinas con capacidad de realizar funciones que realizadas por personas requieren de inteligencia”. Esta definición presenta ciertas deficiencias, ya que se refiere solo a las máquinas y se hace caso omiso de la palabra "artificial". Con todo, el término más problemático en la definición es el de "inteligencia", de forma que dicha definición proporciona una respuesta insuficiente:

una máquina inteligente hace lo mismo que lleva a cabo una persona inteligente. En efecto, esta definición no aclara qué es la inteligencia, sino que remite a aquello que realizan las personas cuando hacen algo inteligente. Partiendo de los supuestos anteriores, hay que tener asimismo en cuenta que en el debate entre los humanistas y los posthumanistas, el plano de las emociones y de los sentimientos se sitúa en la mitad del terreno, es decir, entre el libre albedrío y la conciencia de un lado, y los patrones lógicos del pensamiento racional por el otro. Sea como fuere, lo racional y lo emocional están de tal manera interrelacionados, que se podría sostener que hasta cierto punto son complementarios, y no contradictorios. Las emociones han sido motivo de controversia en el ámbito de la mente y el conocimiento humano. La razón y la emoción han estado enfrentadas, siendo éstas últimas consideradas de forma negativa por mucho tiempo. Actualmente áreas de estudio como la lingüística y la teoría del significado, así como distintas disciplinas científicas, se encargan del análisis de las emociones. Asimismo, la IA y la robótica avanzan considerablemente gracias a los nuevos descubrimientos por parte de la neurociencia en el campo de la inteligencia emocional. Por lo tanto, el término “inteligencia” no puede ser definido en función de los dictámenes de la racionalidad clásica, que no contempla la importancia de las emociones con respecto a la inteligencia, la razón y la toma de decisiones.

4. Conclusión y vías abiertas

En este trabajo se ha tenido en cuenta la problemática que gira en torno a la IA y a la robótica. Para ello, en los antecedentes he querido mostrar una aproximación al debate suscitado sobre la existencia humana entre la postura humanista y la nueva Humanidad Plus o posthumanismo. Comienzo por señalar brevemente el momento de la Ilustración y la importancia que se le dio a la razón para solucionar y mejorar las condiciones materiales y sociales de vida. Fue una idea que logró instaurarse no solo en los pensamientos sobre la naturaleza de la existencia humana, sino también en el mundo interior y privativo del hombre, su propio mundo. Se confía ciegamente en la razón y en la posibilidad del conocimiento científico como un mecanismo objetivo y seguro para articular la sociedad e impulsar su progreso. Sin embargo, el intento de iluminar la vida del hombre haciendo uso de la razón, termina por desembocar en un nuevo género de barbarie que alcanza su auge en la modernidad y se ha incrementado hasta nuestros días. Por este motivo, se defiende la postura de que el progreso tecnocientífico esconde tras de sí un mecanismo de poder y control que será preciso cuestionar para no terminar por aceptar un fuerte determinismo tecnológico encausado por las directrices del método científico. No podemos confiar en que las nuevas capacidades de los sorprendentes descubrimientos tecnológicos, sean las únicas herramientas encargadas de solucionar todos nuestros problemas. Es importante comprender estas cuestiones para hacer frente a los riesgos y retos de este siglo. Sin duda, vivimos en la sociedad del riesgo, por ejemplo la alteración genética de las semillas hace que nos exponamos a una situación de riesgo con el simple hecho de alimentarnos, el mismo peligro ocurre cuando el mercado capitalista arrasa con los recursos naturales para alcanzar un máximo beneficio y provoca desastres ambientales; o el riesgo de la guerra de los recursos; hay riesgo también en el mundo de la vida, cuando por ejemplo asumimos el compromiso de una deuda hipotecaria.

Cabe preguntarse cuál será la tarea de la filosofía para abordar la cuestión sobre si la ciencia y la tecnología pueden mejorar la calidad de vida en el planeta, o bien ocasionar todo lo contrario. Esta cuestión tiene tras de sí un importante contenido filosófico que implica a su vez las grandes preguntas clásicas de la filosofía: qué podemos saber; qué podemos hacer; y qué cabe esperar. Esto nos lleva a plantearnos hacia dónde se dirige la tecnología de la IA y de la robótica y cuales son también sus riesgos y peligros; cuál será la capacidad ética de este nuevo desarrollo científico y tecnológico; cuales son los métodos y las pautas que deben

contemplar los científicos para que el poder tecnológico no caiga en manos de los fines más macabros; y la cuestión sobre si la humanidad está preparada para manejar estas tecnologías. Como tema que valdría la pena explorar pero que no ha habido ocasión de hacerlo en este trabajo, señalo las investigaciones realizadas por la Oxford Martin School (por ejemplo los estudios sobre la ética y el impacto social del mejoramiento cognitivo y emocional, cuestiones de riesgos e incertidumbres, entre otros). En definitiva este instituto se encarga de muchas cuestiones ligadas al transhumanismo y la condición humana, que no se han podido profundizar en el trabajo. También ha quedado pendiente el tema de la identidad, dicho en otras palabras, si ganamos o perdemos identidad con las nuevas tecnologías. La tecnología de la IA y de la robótica avanzada no solo puede ser entendida como una herramienta capaz de reproducir o traspasar los límites biológicos, sino que además supone una programación de la vida, del deseo y de la sociabilidad.

Bibliografía

- **BOSTROM, N. (2016):** *superinteligencia. Caminos, peligros y estrategias.* Teell.
- **CORTINA, A. Y SIERRA, M. (2015):** *¿Humanos o posthumanos? Singularidad tecnológica y mejoramiento humano.* Fragmenta.
- **DE CÓZAR, J. (2002):** *Tecnología, civilización y barbarie.* Anthropos.
- **GÓMEZ, A. Y CANALES, A. (2009):** *Ciencia y Fascismo.* Laertes.
- **HARARI, Y. (2016):** *Homo Deus. Breve historia del mañana.* Penguin Random House Grupo Editorial.
- **HEIDEGGER, M. (2000):** *Ontología: Hermenéutica de la facticidad.* Alianza
- **HORKHEIMER, M. ADORNO, T. W. (2009):** *Dialéctica de la Ilustración. Fragmentos filosóficos.* Ed. Trotta.
- **HORKHEIMER, M. (2002):** *Crítica de la razón instrumental.* Trotta.
- **HORKHEIMER, M. (2000):** *Teoría tradicional y teoría crítica.* Paidós Ibérica
- **KURZWEIL, R. (1994):** *Máquinas inteligentes.* Conacyt.
- **MUMFORD, L. (2011):** *El pentágono del poder. El mito de la máquina (dos).* Pepitas de calabaza.
- **MUMFORD, L. (2010):** *El mito de la máquina. Técnica y evolución humana.* Pepitas de calabaza.
- **NIETZSCHE, F. (2013):** *El origen de la Tragedia.* Austral.

- **NIETZSCHE, F. (2013):** *La genealogía de la moral*, Alianza editorial.
- **SLOTERDIJK, P. (2000):** *Normas para el parque humano*. Siruela.