

LOS DESAFÍOS ÉTICOS DEL TRANSHUMANISMO

ADELA CORTINA
Universidad de Valencia

RESUMEN: Sin duda la revolución 4.0 es ya el mundo en el que vivimos y proporciona instrumentos muy valiosos para construir una sociedad que pueda servir mejor a los seres humanos y al cuidado de la naturaleza desde las nuevas posibilidades. Pero para emprender ese camino es preciso optar por una ética que reconozca que lleva la intersubjetividad entera en su seno, porque es consciente de que no somos individuos aislados, sino que devenimos personas desde el reconocimiento recíproco como interlocutores válidos, como seres dotados de dignidad y no un simple precio, como quienes son carne de la misma carne y hueso del mismo hueso. Ésta sería una ética dialógica de la razón cordial.

PALABRAS CLAVE: crítica de la razón; crítica de la ideología; transhumanismo tecnocientífico; transhumanismo cibernético; ética dialógica; razón cordial.

The ethical challenges of transhumanism

ABSTRACT: Undoubtedly the 4.0 revolution is already the world in which we live and provides very valuable instruments to build a society that can better serve human beings and care for nature from new possibilities. But in order to undertake this path, it is necessary to opt for an ethic that recognizes that it carries intersubjectivity embedded in it, because it is aware that we are not isolated individuals, but that we become people from the reciprocal recognition as valid interlocutors, as beings endowed with dignity and not a simple price, like those who are meat of the same meat and bone of the same bone. This would be a dialogic ethic of cordial reason.

KEY WORDS: Critique of Reason; Critique of Ideology; Techno-Scientific Transhumanism; Cybernetic Transhumanism; Dialogic Ethics; Cordial Reason.

1. LA NECESIDAD DE LA CRÍTICA DE LA RAZÓN Y DE LA CRÍTICA DE LA IDEOLOGÍA

Como es sabido, Kant permaneció largo tiempo sin publicar ningún libro porque se enfrentó al grave problema de las antinomias de la razón¹. Se producen las antinomias cuando se aventuran afirmaciones de *totalidad* que entran entre sí en contradicción y no existe ninguna piedra de toque en la experiencia que permita discernir si son verdaderas o falsas. En su caso, menciona cuatro referidas al espacio, el tiempo, a la contraposición entre determinismo y libertad y al problema de decidir si el mundo tiene una causa. La imposibilidad de resolver las antinomias tiene un efecto muy grave, porque lleva a dudar de la razón como facultad preparada para proporcionar conocimiento objetivo en los asuntos que más preocupan a la humanidad. Precisamente por eso se hace necesario recurrir a la crítica de la razón.

¹ Este estudio se inserta en el Proyecto de Investigación Científica y Desarrollo Tecnológico FFI2016-76753-C2-1-P, financiado por el Ministerio de Economía y Competitividad (ahora Ministerio de Ciencia, Innovación y Universidades), y en las actividades del grupo de investigación de excelencia PROMETEO/2018/121 de la Generalidad Valenciana.

Con la crítica —dirá Kant desde el prólogo a la segunda edición de la *Crítica de la razón pura*— «es posible cortar las mismas raíces del materialismo, del fatalismo, del ateísmo, de la incredulidad librepensadora, del fanatismo y la superstición, todos los cuales pueden ser nocivos en general, pero también las del idealismo y del escepticismo, que son más peligrosos para las escuelas y que difícilmente pueden llegar a las masas»².

Sin una crítica que permita averiguar hasta dónde pueden llegar nuestras facultades cognoscitivas pretendiendo proporcionar conocimiento objetivo, se produce una *metábasis eís állo génos*, el traspaso a otro ámbito de la realidad, que no es aquel en el que un concepto se ha generado. Es lo que ocurre en las antinomias, cuando se habla, por ejemplo, de las relaciones entre causa y efecto que pueden ser empíricamente comprobables, y en un momento determinado el autor abandona el campo de la experiencia y busca para un efecto una causa que se encuentra más allá de la experiencia, actual o posible. Es el caso de Newton cuando recurre a Dios para explicar el espacio como *sensorium Dei*. La afirmación entonces no puede pretender ser considerada como científica, el límite de las ciencias es la experiencia posible, y una afirmación que rebasa el límite de la experiencia puede ser verdadera o falsa, pero en cualquier caso no es científica, porque no hay modo científico de comprobación. Quien la sigue manteniendo como si fuera científica está actuando de forma dogmática, sitúa su presunto saber a resguardo de cualquier crítica, la inmuniza frente a la crítica.

Kant era un profesor de filosofía, que creía necesario llevar adelante el proceso de ilustración de la humanidad, en el que los científicos desempeñarían un papel decisivo, pero siempre que fueran conscientes de los límites de las ciencias (física, biología, geografía, historia) y no cayeran en la tentación de hacer afirmaciones de totalidad sobre las que cabe formular opiniones muy diversas, pero es imposible encontrar una piedra de toque en la experiencia actual o posible para discernir. Evidentemente, es lícito aventurar bosquejos de totalidad y de futuro como *idea regulativa* desde los avances con los que ya se cuenta, pero no darlos por realidad contrastada. Podríamos decir, con el lenguaje médico actual, «fundada en la evidencia».

Ciertamente, existen diversas modalidades del Transhumanismo y del Posthumanismo, objeto de este congreso, pero las más atractivas para el público son las más radicales, las que han podido caracterizarse como una auténtica *utopía*, como una cosmovisión de totalidad que amalgama tesis filosóficas, científicas, tecnológicas y sociales, teniendo como base sobre todo el progreso exponencial de las tecnociencias (Ferry, 2017; Diéguez, 2017). Asegurar la muerte de la muerte (Cordeiro/Wood, 2018), tratar el envejecimiento como una enfermedad que puede y debe ser curada, pronosticar el advenimiento de superinteligencias en la línea de la singularidad, y todo ello en un plazo determinado, que pueden conocer las actuales generaciones, hasta el punto de

² KANT, 1956, B XXXIV, 1978, p. 29

que Kurzweil asegure que en 2045 se conseguirá la singularidad tecnológica, pueden, en efecto, ser afirmaciones utópicas, pero no científicas. Y justamente Transhumanismo y Posthumanismo cifran su credibilidad, pretenden infundir confianza en el público porque dicen basarse en las tecnociencias³.

Sin embargo, el prestigio de la ciencia está ligado a su capacidad de verificación o falsación, dentro del marco de la experiencia posible. Pretender ganar ese prestigio y credibilidad con afirmaciones que exceden con mucho la posibilidad de contrastación actual o posible, pretendiendo que son científicas, es un engaño palmario, que atenta contra la meta de la ciencia, contra su *êthos* más básico. Una utopía científica es simplemente un engaño.

El *primer desafío ético* que plantean el Transhumanismo y el Posthumanismo radicales consiste en dilucidar si sus propuestas de totalidad —sus utopías— tienen valor científico, que es sobre el que descansa su crédito. Es decir, qué hay de científico y qué de *metábasis eís állo génos*. Evidentemente, es real y posible la prolongación de la vida⁴, el aplazamiento de la muerte por enfermedad, la mejora de la calidad de vida en la tercera y cuarta edades, la creación de inteligencias ampliadas, la construcción de inteligencia artificial, la biomejora de capacidades humanas y todas las ventajas que comporta la industria 4.0. Pero el salto a la inmortalidad, la salud perpetua y las superinteligencias, incluso a la inteligencia artificial general, no es científico.

Y este salto ilegítimo provoca sospechas que exigen *practicar la crítica en un segundo sentido*, relacionado ahora no con Kant, sino con Marx y la Teoría Crítica de la Escuela de Frankfurt. Como bien decía Habermas, la ciencia y la técnica se pueden convertir en ideología, en esa visión deformada y deformante de la realidad que permite satisfacer los intereses espurios de los poderosos para practicar la dominación. En este caso, los intereses de industrias como la farmacéutica y plataformas potentes, que consiguen obtener una gran cantidad de recursos, utilizando la opinión pública y las redes sociales como medio de ideologización.

En efecto, en 1968 publica Habermas su célebre artículo «Ciencia y técnica como “ideología”», en el que, en diálogo con Marcuse, muestra cómo la ciencia y la técnica que debían servir como fuerzas productivas, según el esquema marxiano, se han convertido en parte de la estructura ideológica que legitima la dominación (Habermas, 1984a). Como puntualizarán tanto Habermas como Apel, es imprescindible delimitar los roles de los dos intereses que lideran el progreso del conocimiento en la evolución de la especie —el científico-técnico de dominación y el práctico en la comprensión entre los sujetos— de modo que sea el diálogo entre los sujetos el que decida sobre el uso de la ciencia y de la técnica

³ Las nuevas tecnologías convergen en el acrónimo NBIC. Es decir, Nanotecnología, Biotecnologías, Informática (big data, internet de las cosas) y ciencias cognitivas (IA, robótica, nuevas técnicas de hibridación, impresoras 3D (FERRY, 2017, p. 14).

⁴ En este sentido, un desafío ético inminente consiste en preguntar cómo organizar esa vida prolongada, que cambia muchas de las claves de nuestras expectativas actuales. Muy interesante en este sentido es el libro de Lynda Gratton & Andrew Scott *La vida de cien años*.

con vistas a la emancipación, con vistas a poner fin a las sociedades ideologizadas (Habermas, 1984b; Apel, 1985; Conill, 2006). Si la ciencia y la técnica —hoy tecnociencias— se convierten en ideología, y dominan con promesas irrealizables el espacio de la opinión pública en que los sujetos deliberan y debaten, entonces sirven para fortalecer la dominación apoderándose del espacio público. Aunque los defensores del Transhumanismo y del Posthumanismo promuevan los diálogos en la opinión pública, si la información del público es falsa, el discurso transhumanista busca legitimar la ideología.

Un segundo reto ético planteado por el Transhumanismo y el Posthumanismo consiste, pues, en analizarlos desde el punto de vista de la crítica de la ideología, dilucidando si no están seduciendo con promesas incumplibles a las pulsiones humanas profundas desde intereses económicos y de poder que recurren al discurso de las utopías científicas. Distinguir entre una comunicación distorsionada, es decir, ideológica, que encubre intereses espurios de dominación, y una comunicación que busca el entendimiento, no la adhesión desinformada, se hace necesario. Es la intersubjetividad humana la que debe liderar el uso de las técnicas, hoy en día las aportaciones tecnocientíficas, desde un diálogo bien informado.

Y un tercer desafío ético consiste en esclarecer a qué tipo de ética se apela para mostrar la bondad moral de las propuestas: ¿a una ética individualista, propia de un liberalismo no social, que adopta la jerga de lo políticamente correcto, aunque en realidad no es coherente con sus propuestas, o a una ética basada en la convicción de que los seres humanos se hacen desde el reconocimiento recíproco, es decir, desde el reconocimiento intersubjetivo, de modo que el progreso tiene que atender a las personas y a los vínculos que les permiten serlo?. El progreso tiene que pasar entonces por adensar los vínculos de una intersubjetividad no ideologizada. Y no puede consistir en perseguir una *utopía*, sino una *meta viable y deseable*. Ésta sería la ética propia de las tecnociencias, que ha de tener en cuenta el quehacerse dialógico de los seres humanos, de los miembros de la familia humana.

Para ir abordando estos asuntos caracterizaremos brevemente el Transhumanismo. Para definirlo resulta de gran utilidad recurrir a la caracterización de quien acuñó el término, el biólogo Julian Huxley, hermano de Aldous Huxley, quien fue el primer director general de UNESCO. En efecto, en *Religion without Revelation* (1927) escribe: «La especie humana puede trascenderse a sí misma, si lo desea, trascenderse a sí misma —no sólo esporádicamente, un individuo aquí de un modo, otro allá de otro modo—, sino en su totalidad, como humanidad. Necesitamos un nombre para esta nueva creencia. Tal vez *transhumanismo* pueda servir: el hombre permaneciendo hombre, pero trascendiéndose a sí mismo, al actualizar nuevas posibilidades de y para su naturaleza humana».

El transhumanismo se distanciaría de una posición como la de Nietzsche, que pretende el autotrascendimiento de algunos individuos con capacidad y voluntad de hacerlo, porque los transhumanistas se proponen como objetivo el autotrascendimiento de toda la humanidad, y además no sólo tratando de

encarnar la fórmula «llega a ser el que eres» de Píndaro y Nietzsche, sino «llega a ser más de lo que eres».

En cualquier caso, los transhumanistas han tenido y están teniendo buen cuidado en distanciarse de anteriores proyectos de modificar a la especie humana de cuño totalitario, levantando la bandera progresista y ligando sus propuestas cada vez más al liberalismo cultural, la democracia política y al igualitarismo. Es el caso, entre otros, de James Hugues, quien considera que la biopolítica está emergiendo como una nueva dimensión de la opinión política. En *Citizen Cyborg* propone un «transhumanismo democrático», que articula la biopolítica transhumanista con la política social democrática y económica y con la política liberal cultural. Entiende que conseguiremos el mejor futuro posthumano cuando aseguremos que las tecnologías son seguras, accesibles a todos y se respeten los derechos individuales en el control de los propios cuerpos. Los beneficios han de llegar a todos, y no sólo a una élite, y en esto el Estado debe intervenir. Sin embargo, ¿pertenece estas afirmaciones al «se dice» de lo políticamente correcto, o son propuestas éticamente fundadas?

En cualquier caso, para lo que ahora pretendemos podemos caracterizar el Transhumanismo de un modo más sencillo como el «intento de transformar sustancialmente a los seres humanos mediante la aplicación directa de la tecnología» (Diéguez, 40). Sea mediante la fusión con las máquinas o alojando nuestras mentes en máquinas, sea manipulando los genes en la línea germinal.

Ciertamente, existen diversos modelos de Transhumanismo atendiendo a diversos criterios de especificación, pero dos modelos son esenciales:

- 1) El Tecnocientífico o Cibernético, inspirado en los trabajos de IA, la ingeniería del software y la robótica (Minsky, Moravec, Kurzweil, Bostrom⁵, Sandberg). Buscan una hibridación del hombre y la máquina, que recurre a la IA y a la robótica, más que a la biología. Tratan de tecnofabricar una posthumanidad, una nueva especie hibridada con máquinas dotadas de capacidades físicas y una IA superior a la humana (Ferry, 40). Es un posthumanismo, una hibridación entre humano y máquina, superación de la inteligencia humana por una artificial, que lleva a una especie nueva que permitirá sustituir a los humanos.
- 2) Transhumanismo biológico, que lleva a la humanidad a su más alto desarrollo; defiende el biomejoramiento humano con base biológica, médica, farmacológica y genética (Harris, Savulescu, Church). La aparición de la biología sintética (creación de genes para fines específicos) ha sido un gran refuerzo porque parece permitir liderar la evolución. La biología sintética es una tecnociencia capaz de transformar la naturaleza, de recrear la vida y rediseñarla de acuerdo con intereses y necesidades. Se iría generando una nueva especie biológica, construida a la medida de nuestros deseos (Diéguez, 112 y ss.).

⁵ En *Superinteligencia*, publicado en 2011 y traducido en 2016, Bostrom se aleja de las posturas transhumanistas que había mantenido (AMOR PAN, 2019, 38).

Ciertamente, cada uno de estos dos modelos plantea problemas éticos específicos, y en este trabajo voy a referirme únicamente al primero de estos dos modelos por razón de espacio y sobre todo porque de la biomejora ya me he ocupado en otros lugares, a los que aquí me permito remitir (Cortina 2013 y 2017, cap. 6).

2. DESAFÍOS ÉTICOS DEL TRANSHUMANISMO TECNOCIENTÍFICO O CIBERNÉTICO

Para abordar los desafíos éticos que plantea el Transhumanismo cibernético es necesario aclarar qué se entiende por «inteligencia» y considerar al menos tres tipos de Inteligencia Artificial (IA).

En principio, conviene recordar que es preciso hablar de «inteligencias», y no sólo de «inteligencia», en el caso de los humanos, en el de los animales, en el de los vegetales, e incluso en el de la tierra según la hipótesis Gaia. Podemos llamar entonces «inteligencia» en sentido amplio a la capacidad de perseguir metas, planificar, prever consecuencias de las acciones y emplear herramientas para alcanzar las metas. La inteligencia sería la capacidad de resolver problemas con instrumentos.

En cuanto a la *inteligencia artificial*, nace en 1955, en un congreso en Los Ángeles sobre máquinas que aprenden. John McCarthy introduce la expresión «inteligencia artificial» en 1956 y se refiere con ella a la creación de máquinas que pueden tenerse por inteligentes porque interactúan con los seres humanos hasta el punto de que una persona ya no sabe si está hablando con una máquina o con otra persona humana. Es lo que recibe el nombre de «test de Turing»⁶.

Pero, por decirlo con mayor precisión, podemos recurrir a la caracterización del *High-Level Expert Group on Artificial Intelligence*, creado por la Comisión Europea, en sus Orientaciones Éticas para una IA confiable, publicadas en 2019, tras haber sacado a la luz y discutido con expertos dos borradores⁷. Según el texto de las *Guidelines*, los sistemas de IA son sistemas de software (y posiblemente también de hardware), diseñados por humanos que, dada una meta compleja, actúan en la dimensión física o digital percibiendo su entorno mediante la adquisición de datos, interpretando los datos recogidos, estructurados o no estructurados, razonando sobre el conocimiento o procesando la información derivada de estos datos y decidiendo las mejores acciones que hay que realizar para alcanzar la meta. Los sistemas IA pueden utilizar reglas simbólicas o aprender un modelo numérico, y pueden también adaptar su conducta analizando cómo el entorno es afectado por las acciones previas.

⁶ Como es bien conocido, TURING (1912-54) es el padre de la informática, gracias a su máquina (1936), que es una visión de IA. En 1950 defiende en *Mind* que los ordenadores pueden tener comportamientos inteligentes.

⁷ *Ethics Guidelines for Trustworthy AI*, abril de 2019, p. 36.

Por otra parte, como disciplina científica, la IA incluye varios enfoques y técnicas, tales como *machine learning* (del que son ejemplos el *deep learning* y el *reinforcement learning*), el *machine reasoning* (que incluye planificar, programar, representaciones de conocimiento y razonamiento, búsqueda y optimización), la robótica (que incluye control, percepción, sensores y actuadores (*actuators*), y la integración de todas las demás técnicas en los sistemas ciberfísicos.

En este ámbito de la inteligencia artificial pueden distinguirse tres modalidades que, a mi juicio, plantean problemas éticos diferenciados:

1) *La inteligencia superior o superinteligencia*. Con esta expresión nos referimos a un tipo de inteligencia que podría superar a la humana, de modo que las máquinas podrían sustituir al hombre. Esta modalidad de IA es la que da lugar a las propuestas transhumanistas y posthumanistas con la idea de la «singularidad».

John von Neumann fue uno de los primeros en vislumbrar la potencia de cálculo que un ordenador podía tener, y llega a afirmar que el progreso en la tecnología y los cambios en nuestra forma de vida «muestran signos de aproximarse a una especie de singularidad esencial en la historia de la especie». En 1993 Vinge escribe *Technological Singularity*, donde propone la superación de la mente humana por máquinas con inteligencia artificial. En 1995 Jastrow propone transferir la mente a la máquina para alcanzar la inmortalidad. Pero la figura más conocida es Raymond Kurzweil, quien recoge la idea en *The Singularity Is Near: When Humans Transcends Biology*. Según él, los humanos dejarán su soporte biológico y pasarán su inteligencia a las máquinas. Esto da lugar a la *Singularity University*, fundada en 2008 con el patrocinio de Google y NASA, cuyo lema es: preparando a la humanidad para un cambio acelerado de tecnología.

Habría entonces un cambio de sustrato entre inteligencia humana e inteligencia artificial: la Singularidad necesita sobrepasar los límites impuestos por el tejido nervioso y el sustrato de la inteligencia artificial sería de silicio. Los seres humanos seríamos un elemento más en la cadena de la evolución que culminará en esos seres singulares. No se trataría en modo alguno del superhombre nietzscheano, para el que el cuerpo es esencial (Jesús Conill, 2019), sino de seres singulares en los que el cuerpo biológico ha sido sustituido por la máquina. Sería una especie nueva.

Sin duda, existen amplias discrepancias en el ámbito de la IA sobre si estos pronósticos del «transhumanismo» y del «posthumanismo» van a cumplirse por tener base científica suficiente para ello. Algunos autores dan por sentado que se llegará a crear superinteligencias artificiales en este mismo siglo; en concreto, Kurzweil considera que en 2045 se conseguirá la singularidad tecnológica, gracias al incremento exponencial de las tecnologías de la información (Kurzweil, 2005/2012). Sería posible reubicar nuestro cerebro en un soporte no biológico y garantizar su perdurabilidad *ad infinitum*. Según José Luis Cordeiro y David Wood veremos la muerte de la muerte en 2045 Cordeiro/Wood, 2018). Sin embargo, otros entienden que no existe

base científica para esa suposición, cuando en realidad sabemos tan poco del cerebro (Boden, 2017, López de Mántaras, y Meseguer, 2017, 13; Diéguez, 2017).

Pero la sola hipótesis ya abre un mundo de cuestiones éticas, que es preciso abordar porque los transhumanistas afirman que es un deber moral trabajar en la línea de intentar trascender la especie humana con todas sus imperfecciones para crear esos seres perfectos que compondrían la singularidad. Si el ser humano es intrínsecamente imperfecto, entonces, según ellos, es un deber moral buscar su mejoramiento por medios técnicos. Sin embargo, dejando de lado otras cuestiones, la pregunta se impone: ¿es realmente un deber moral intentar trascender la especie humana en una presuntamente superior, en realidad desconocida, dividiendo el mundo de la inteligencia racional en dos especies, una superior y otra inferior? ¿No estaríamos abonando un mundo de amos y esclavos, en el que los segundos no tendrían la menor posibilidad de hacer una revolución, sino que estarían a merced de las superinteligencias? ¿Serían ésta la culminación del proceso de ilustración de la humanidad?

El Salvaje de *Un mundo feliz* de Huxley se encontraba en una mejor posición, porque tenía libertad para elegir entre el mundo feliz y el humano, aunque éste llevara aparejados el sufrimiento, el envejecimiento y la muerte, pero en el mundo de las superinteligencias la especie de los esclavos no tendría posibilidad de elección. ¿Es realmente un imperativo ético construir ese mundo?

Por otra parte, ¿cuál será la ética de esas superinteligencias? Nick Bostrom, uno de los adalides del posthumanismo, al menos en su primera etapa, aconseja integrar valores en esas inteligencias que, aprendiendo, se independizarán de los humanos (Bostrom, 2014). Pero —a mi juicio— si esto fuera posible, y las máquinas aprendieran por su cuenta, poco podríamos hacer por conseguir que siguieran manteniendo como valores el respeto, la solidaridad, la justicia o la compasión. Serían los propios sistemas superinteligentes los que irían proponiendo sus valores y actuando o no de acuerdo con ellos. Ésta sí que sería una «ética de la inteligencia artificial», que no estaría en nuestras manos. ¿Es ahora un deber moral propiciarla?

Feuerbach criticaba la religión asegurando que los hombres pobres crean un Dios rico, que los convierte en objeto de su amor. Pero esas superinteligencias, de las que nada sabemos, ¿por qué habrían de respetar a los seres humanos? A mi juicio, nunca es un imperativo ético, ni científico, construir sociedades enigmáticas e irreversibles, incapaces de infundir confianza.

En tercer lugar, no existe hasta el momento ninguna evidencia de que vaya a ser posible en ningún momento crear esas superinteligencias, que no necesitan contar con un cuerpo biológico. ¿La promesa, no científica de lograrlo, no es un señuelo para conseguir financiación, prestigio y poder para sus promotores?

Y sobre todo, en un mundo en que es una realidad sangrante el sufrimiento causado por las guerras, la pobreza, la aporofobia y la injusticia, ¿es un

deber moral invertir una ingente cantidad de recursos en construir presuntos seres pluscuamperfectos, o es el modo en que empresas poderosas consiguen todavía más riqueza y poder? ¿No es una exigencia ética palmaria utilizar los grandes beneficios de la inteligencia artificial para resolver estos problemas acuciantes?

2) Un segundo tipo de inteligencia es la *inteligencia general*, aquella que puede resolver problemas generales. Ésta es la forma de inteligencia típicamente humana, y constituye el fundamento de la IA, en que trabajan las mentes más brillantes: el objetivo de la IA, como disciplina científica, es conseguir que una máquina tenga una inteligencia de tipo general, similar a la humana.

Fue John Searle quien distinguió ya en 1980 entre IA fuerte y débil (Searle, 1980). La fuerte implicaría que un ordenador es una mente y es capaz de pensar igual que un ser humano, pero lo que intenta demostrar Searle es que la IA fuerte es imposible, porque la máquina carece de la *intencionalidad* por la que los humanos damos significado a lo que nos rodea: una máquina no conoce el significado de los símbolos que maneja. *Sin un cuerpo* las representaciones abstractas carecen de contenido semántico: no puede haber inteligencia general sin cuerpo⁸.

Este punto es central: las máquinas carecen del conocimiento de *sentido común* que es posible por nuestras vivencias corporales. El cuerpo es esencial para dar significado a lo que nos rodea mediante la intencionalidad, para comprender e interpretar desde los contextos concretos, para contar con valores, emociones y sentimientos, para tomar decisiones desde ese *êthos*. La cuestión es entonces: ¿es posible dotar de sentido común a las máquinas, aunque no tengan un cuerpo como el humano? ¿Sería posible pertrecharles de conciencia moral, autonomía y responsabilidad? Realmente, la financiación que reciben quienes trabajan en ello es astronómica, pero por el momento no parece haberse logrado⁹.

Sin embargo, en el caso de que fuera posible construir sistemas inteligentes con una inteligencia general como la humana, ¿tendríamos que aceptar que están dotados de autonomía y, por lo tanto, son personas y que, en consecuencia, es preciso reconocerles dignidad y exigirles responsabilidad?, ¿tendrían derechos y deberes?, ¿deberíamos tratarles con respeto y compasión?, ¿deberían ser ciudadanos del mundo político, elegibles como representantes en sociedades democráticas, sin estar manejadas por un ser humano? ¿Se forjarían a través del reconocimiento recíproco?

Si fuera posible, sería así, y correríamos riesgos como que los fabricantes

⁸ A este respecto es célebre el experimento de la caja china del que habla Searle: una manipulación sintáctica sin comprensión semántica carece de sentido.

⁹ «Hasta ahora, sin embargo, los programadores humanos han sido incapaces de idear algoritmos capaces de capturar y expresar profundas servas de “sentido común” y *savoir faire* et *pouvoir faire* de los humanos (KEANE, 2019, p. 30).

de tales máquinas podrían crear ejércitos que acabarían tomando decisiones en política, en economía y en todos los ámbitos de la vida social. Pero parece sumamente improbable que se consiga, o más bien imposible, y no sólo porque no se ha conseguido, sino también porque simularían intencionalidad, emociones, valores y sentido común, pero no dejaría de ser una simulación. Harían «como si» sintieran, pero para sentir se necesita un cuerpo Ferry, 2017, pp. 204 y 205; López de Mántaras y Meseguer, 2017).

3) Por último, *la inteligencia especial* es la que lleva a cabo trabajos específicos, es la propia de sistemas inteligentes capaces de realizar tareas concretas de forma muy superior a la inteligencia humana, porque pueden contar con una inmensa cantidad de datos y también con algoritmos sofisticados, que pueden llevar a resultados. Es lo que tenemos en diversos ámbitos.

El caso más conocido es el de la supercomputadora de IBM *Deep Blue*, que jugó al ajedrez con Gary Kasparov, campeón del mundo, en 1996 y 1997. En 1996 ganó Kasparov, pero en 1997, *Deep Blue* había aprendido de sus errores y derrotó a Kasparov. El revuelo fue enorme. Pero es que el sistema inteligente puede llevar a cabo tareas concretas contando con una infinidad de datos y con una capacidad de correlación muy superior a la de los seres humanos.

En este sentido, hay una gran cantidad de problemas que están siendo abordados con ayuda de buenos algoritmos, en el sector de la salud (analizar los síntomas de un paciente en muy distintas modalidades de la medicina, hacer un diagnóstico y proponer un tratamiento), en la predicción climatológica, en la productividad y eficiencia empresarial, en la comunicación, el ocio, la planificación del tiempo, el ahorro de tiempo, el abaratamiento de costes, en el asesoramiento a la hora de conceder un crédito, reconocer voces humanas y leer textos, aconsejar en el ámbito agrícola. El proceso consiste en construir un patrón, que permita adivinar el comportamiento futuro, porque se dice que somos humanos predecibles. La búsqueda sistemática de un patrón en un amplio registro histórico se llama *minería de datos* (*data mining*) y se utiliza de forma rutinaria tanto en investigación científica como en el mundo de los negocios (Latorre, pp. 91-95).

Sin embargo, en todos estos casos el elemento directivo sigue siendo la persona humana que se vale de la potencia del sistema inteligente para calcular y tratar gran cantidad de datos, incluso para aprender de sus «experiencias».

Es en este tipo de IA en el que actualmente nos encontramos y plantea una gran cantidad de desafíos éticos de los que es preciso ocuparse: cómo organizar el mercado laboral para que puedan protegerse los derechos sociales de toda la ciudadanía, teniendo en cuenta que va a cambiar totalmente; cómo proteger la autonomía de las personas frente a las máquinas autómatas, que no autónomas; cómo organizar un mundo en que los algoritmos deciden gran parte de nuestras vidas de modo que quede a salvo nuestra autonomía.

A mi juicio, preocuparse de cómo aprovechar las ventajas de la revolución 4.0 de modo que beneficien a las personas, autónomas y vulnerables y no les

dañen, es el gran imperativo ético relacionado con la IA. Es una realidad que hay que enfocar de forma deseable¹⁰.

3. ¿DESDE QUÉ ÉTICA?

Desde el comienzo de esta intervención entendimos que el primer desafío ético en el ámbito del Transhumanismo y el Posthumanismo consiste en distinguir entre las propuestas que están científicamente avaladas por tener un respaldo en la experiencia, actual o posible, y aquellas otras que no gozan de un respaldo científico. Poner a la luz abiertamente que es científicamente ilegítimo el tránsito del primer tipo de propuestas a las segundas, tratándolas a todas como si tuvieran un aval científico, es lo primero que exige una ética de la ciencia, entendida como ética de la responsabilidad. Se trata con ello de practicar la crítica de la razón, que es el camino ilustrado hacia un saber confiable. Teniendo en cuenta que la confianza es indispensable para el funcionamiento de las sociedades, sea en las tecnociencias, en la política, en la economía o en las demás actividades de la vida compartida.

Un segundo reto ético consistiría en tratar de desentrañar si llevar a la opinión pública determinados proyectos posthumanistas y transhumanistas como si estuvieran bien acreditados no es en realidad hacer un uso ideológico de las tecnociencias, dado el poder económico y político que entrañan. Practicar la crítica de la ideología se haría necesario para evitar que sea el interés técnico por dominar, en manos de poderes económicos y políticos, el que lleve las riendas del presente y el futuro. Cuando lo bien cierto es que, desde un punto de vista ético, son los sujetos afectados por las consecuencias de las tecnociencias, es decir, las personas, las que tienen que decidir el futuro desde el diálogo abierto y contando con información fidedigna. Es el interés práctico en el entendimiento mutuo y el acuerdo el que tiene que poner también a las tecnociencias en el camino de construir una sociedad emancipada, libre de ideologías.

Porque sin duda la revolución 4.0 es ya el mundo en el que vivimos y proporciona instrumentos muy valiosos para construir una sociedad que pueda servir mejor a los seres humanos y al cuidado de la naturaleza desde las nuevas posibilidades. Pero para emprender ese camino es preciso optar por una ética que reconozca que lleva la intersubjetividad entrañada en su seno, porque es consciente de que no somos individuos aislados, sino que devenimos personas desde el reconocimiento recíproco como interlocutores válidos, como seres dotados de dignidad y no un simple precio, como quienes son carne de la misma carne y hueso del mismo hueso. Ésta sería una ética dialógica de la razón cordial (Cortina, 2007, 2017).

¹⁰ Me he ocupado de la ética de la inteligencia artificial especial en la intervención en la Real Academia de Ciencias Morales y Políticas del 7 de mayo de 2019 con el título «Ética de la inteligencia artificial».

REFERENCIAS BIBLIOGRÁFICAS

- AI4People, *Ethical Framework for a Good AI Society: Opportunities, Risks, Principles and Recommendations*, *Minds and Machines* (2018), 28.689-707.
- Amor Pan, J. R. (2018). «Mejora humana, posthumanismo, liberalismo y capitalismo: ¿Los cuatro jinetes del capitalismo?», en Domingo García-Marzá, José Félix Lozano, Emilio Martínez y Juan C. Siurana (eds.), *Ética y Filosofía política*. Madrid: Tecnos, 273-283.
- Amor Pan, J. R. (2019). «El sueño de Gilgames hecho realidad: somos la última generación humana mortal», en *Diálogo filosófico*, n° 103, 27-42.
- Apel, K.-O. (1985). *La transformación de la filosofía*. Madrid: Taurus, I.
- Boden, M. A. (2017). *Inteligencia Artificial*. Madrid: Turner Noema.
- Bostrom, N. (2005). «A History of Transhumanist Thought», *Journal of Evolution and Technology*, vol. 14, 1, 1-30.
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press (trad. cast. Superinteligencia, caminos, peligros, estrategias. Zaragoza: Teell, 2016).
- Conill, J. (2006). *Ética hermenéutica*. Madrid: Tecnos.
- Conill, J. (2019). *Intimidación corporal y persona humana. De Nietzsche a Ortega y Zubiri*. Madrid: Tecnos.
- Cordeiro, J. L. y Wood, D. (2018). *La muerte de la muerte*. Barcelona: Deusto.
- Cortina, A. (2011). *Neuroética y neuropolítica*. Madrid: Tecnos.
- Cortina, A. (2013). «Neuromejora moral: ¿un camino prometedor ante el fracaso de la educación?», publicado en *Anales de la Real Academia de Ciencias Morales y Políticas*, Madrid, Año LXV, n° 90, 313-331.
- Cortina, A. (2007). *Ética de la razón cordial*. Oviedo: Nobel.
- Cortina, A. (2017). *Aporofobia, el rechazo al pobre*. Barcelona: Paidós.
- Diéguez, A. (2017). *Transhumanismo. La búsqueda tecnológica del mejoramiento humano*. Barcelona: Herder.
- Ferry, L. (2017). *La revolución transhumanista. Cómo la tecnomedicina y la uberización del mundo van a transformar nuestras vidas*, Alianza (original: Plon, 2016).
- Feito, L. (2019). «El Transhumanismo», en *Diálogo filosófico*, n° 103, 4-25.
- Gratton, L. & Scott, A. (2018). *La vida de cien años*. Bilbao: Lettera Publicaciones.
- Habermas, J. (1984a). «Ciencia y técnica como “ideología”», en *Ciencia y técnica como «ideología»*. Madrid: Tecnos, 53-112.
- Habermas, J. (1984b). «Conocimiento e interés», en *Ciencia y técnica como «ideología»*. Madrid: Tecnos, 159-181.
- High-Level Expert Group on Artificial Intelligence set up by the European Commission (2019), *Ethics Guidelines for Trustworthy AI*, abril de 2019.
- Hugues, J. (2004). *Citizen Cyborg: why democratic societies must respond to the redesigned human of the future*. Cambridge: Westview Press, MA.
- Huxley, J. (1927). *Religion without Revelation* (1927). Londres: E. Benns.
- Kant, I. (1956). *Kritik der reinen Vernunft*, Felix Meiner Verlag, Hamburg (trad. cast. de Pedro Ribas. Madrid: Alfaguara, 1978).
- Keane, J. (2019). «La nueva era de la revolución de las máquinas», *Letras libres*, abril, año XVIII, 24-32.
- Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcends Biology*, Penguin (hay trad. esp. en Berlin: Lola Books, 2012).
- Latorre, J. I. (2019). *Ética para máquinas*. Barcelona: Ariel.

- Llopis, R. (2019). «La inteligencia artificial como factor de innovación en la empresa», conferencia pronunciada en el Seminario de la Fundación ÉTNOR el 14 de marzo de 2019.
- López de Mántaras, R. y Meseguer, P. (2017). *Inteligencia artificial*. Madrid: Los Libros de la Catarata/CSIC.
- Searle, J. R. (1980). «Minds, Brains and Programs», *Behavioral and Brain Sciences* 3 (3), 417-457.

Universidad de Valencia
Catedrática Emérita de Ética y Filosofía Política
adela.cortina@uv.es

ADELA CORTINA

[Artículo aprobado para publicar en febrero de 2022]

Este artículo responde a la ponencia tenida en las Jornadas sobre *Transhumanismo*, que tuvieron lugar en la Universidad Comillas, Escuela Técnica Superior de Ingeniería, Cátedra Hana y Francisco José Ayala de Ciencia, Tecnología y Religión, los días 29 al 31 de mayo de 2019.