Hacia el *Humanismo Digital* desde un denominador común para la *Cíber Ética* y la *Ética de la Inteligencia Artificial*

JOSÉ LUIS FERNÁNDEZ FERNÁNDEZ

§1. Una *pregunta de investigación* desde la *Filosofía de la Técnica*

La reflexión filosófica acerca de la técnica y de la tecnología no es nueva, en absoluto, en la historia del pensamiento. De hecho, enlaza entre sí aspectos de muy variado tenor, que van, por ejemplo, desde la Antropología (López Moratalla 2017; Beorlegui, 2016; Arana, 2019), hasta las conexiones y alcances que apuntan a la dimensión económica de la vida humana en sociedad. Aquella consideración filosófica de la técnica se prolonga también hacia elementos cercanos al ámbito de la Política, en lo que concierne a la organización y a la estructura de las sociedades y de los grupos humanos estables. Pero sobre todo, una consideración sistemática sobre la realidad de la técnica, se conecta de manera inmediata con la Teoría del Conocimiento; es decir, con la indagación acerca del modo como los seres humanos conocen la realidad y operan sobre ella de manera pragmática, aplicando el saber teórico–práctico objetivado y lingüísticamente disponible. Al fondo de todo ello, sin duda, emerge la cuestión por el sentido —ya implícito, ya expreso— de la actividad tecnológica y la significación de la técnica en el despliegue y en la dinámica de lo humano (Heredia 2018).

Como decimos, en efecto, la tecnología y la técnica, han sido objeto de estudio, han servido de ocasión para llevar a término sutiles análisis acerca del conocer y sus diversos modos; y han dado pie a una sustanciosa reflexión que, en suma, ha contribuido a desvelar algunos de los ámbitos más enigmáticos y misteriosos del espíritu humano, en la inacabada tarea de *hacer* la vida en el marco de un entorno físico–natural que le ofrece posibilidades, al paso que le fija límites, más o menos infranqueables, en función, entre otras cosas, del grado del desarrollo tecnológico (Beorlegui 2018).

J. L. Fernández Fernández (✉)
Universidad Pontificia Comillas, España
e–mail: jlfernandez@icade.comillas.edu

Disputatio. Philosophical Research Bulletin
Vol. 10, No. 17, Jun. 2021, pp. 107–130
ISSN: 2254–0601 | [SP] | **ARTÍCULO**

La naturaleza, el medio en el que —digámoslo así— la mera *biología* humana deviene plena *biografía*, tanto personal cuanto colectiva —desde el grupo base de pertenencia e inserción natural; hasta lo humano, considerado a escala planetaria e imbricado en el cosmos— ofrece recursos para la vida. El hombre, consciente como va siendo cada vez más de lo lábil que son los procesos ecológicos de los que su propia vida depende, está todavía en proceso de aprendizaje para utilizar los recursos del medio con prudencia y para, desde la Ética, administarlos con responsabilidad. Por lo demás, dada la economicidad de aquellos bienes naturales —esto es, escasos, finitos, susceptibles de usos alternativos y, la mayoría de ellos, no renovables— se le plantean a la humanidad también serios retos a los que ésta ha venido tratando de responder de múltiples formas, más o menos eficientes, a lo largo de la historia, mediante un despliegue de imaginación creativa, enraizada en una constante antropológica, cual es la capacidad para externalizarse de forma innovadora.

Llevar a efecto un abordaje reflexivo sobre las concreciones fácticas que representa este proceso de adaptación del entorno para ponerlo al servicio de metas y objetivos humanos, constituye, precisamente, el objeto de estudio de la *Filosofía de la Técnica* (Esquirol 2011). Nosotros, en este trabajo, no vamos a poder entrar a fondo en ello, toda vez que el objetivo que nos mueve no es abordar esa específica tarea; sino otro, más acotado y axiomático. A saber, el de tratar de responder a una pregunta de investigación que sin ser retórica, en absoluto, se formula en el marco de una apuesta inequívoca por lo que cabría denominar la búsqueda de un *Humanismo Digital*. La pregunta que guía nuestra reflexión en este trabajo puede quedar adecuadamente formulada en los siguientes términos:

> *¿Cabría identificar un común denominador ético que pudiera servir de propuesta universal para un* Humanismo Digital*; esto es, una situación en que la humanidad, tornando en favor propio los adelantos técnicos propios de la* Cuarta Revolución Industrial (Schwab 2017)*, al paso que consigue progresivamente minorar penosidades y sufrimientos indeseables y evitables, lograra desplegar sus potencias y capacidades para así acceder, entre otras cosas, a las cotas más altas de desarrollo económico, a los más elevados estándares de progreso técnico y, sobre todo, a la cumbre más prominente del florecimiento humano, en el marco de la* Cíber Sociedad (Stückelberger 2018); *ubicando, por modo de axioma, lo humano y la dignidad de la persona en el centro de todas las aplicaciones tecnológicas conexas con* la Inteligencia Artificial *y su posible desarrollo futuro?*

§2. Triple realidad y ambivalencia del hecho técnico

Aunque, como va dicho, renunciamos expresamente a abundar de manera sistemática en ello, habida cuenta, sin embargo, de que la *Filosofía de la Técnica* constituye el marco en el que se haya de insertar lo que vayamos a decir más abajo en relación con la *Cíber Ética* (Sweeney 2018) y la *Ética de la Inteligencia Artificial* (Coeckelbergh 2021), conviene llevar a efecto unas indicaciones, siquiera sean mínimas, que sirvan de referencia para una mejor comprensión de lo que abordemos en el cuerpo central de este trabajo.

Una primera consideración básica respecto de la técnica debiera consistir en subrayar una triple realidad respecto de aquélla. Pues, en línea con las tesis estructurales propias de la *Sociología del Conocimiento* (Berger y Luckmann 1979), cabría afirmar que la técnica se declina, a la vez, en tres momentos, dialécticamente enlazados entre sí, en una dinámica autopoyética sujeta a un proceso constante de retroalimentación. A saber: una capacidad de obrar; un saber concretado en una serie de datos objetivos, susceptibles de estudio y transmisión; y un poderoso agente de configuración de lo humano en todas las facetas y ámbitos en los que aquello se expresa.

Así, de una parte, la capacidad técnica se manifiesta como una de las dimensiones antropológicas básicas a través de las que se dice lo humano en el mundo —la característica del *homo* como *faber*—. En segundo término, la técnica, objetivamente considerada, en esencia, no es sino un producto humano, que cristaliza como saber; y que constituye una realidad objetiva, empíricamente reconocible como acervo de un conocimiento aplicado que se ha venido acumulando a lo largo de la historia de la humanidad como tecnología. En último extremo, la técnica constituye, a la vez, una relevante vía de construcción de sociedad y de cultura; y que, por consiguiente, está en condiciones de configurar de forma sustancial la vida entera y el completo destino de la humanidad y de lo humano en su conjunto (Cortina y Serra 2016).

Para acabar de delinear el escenario, anotemos cómo los inevitables intereses del conocimiento (Habermas y Husserl 1995) hacen que la técnica nunca pueda resultar neutra y que, incluso, sea frecuente que vaya entrelazada con una elevada carga ideológica (Habermas 1985) que la sitúa —explícita o de forma tácita— al servicio de unos objetivos y por referencia a unas metas que, en todo caso, habrían de ser revisadas con talante crítico.

Por lo demás, como hemos ya dado a entender, tomar la técnica como objeto de estudio filosófico supone, de una parte, conectar con intuiciones formuladas a lo largo de una amplia tradición en la historia del pensamiento,

cuyos hitos principales cabría ubicar por referencia a los tres momentos siguientes: de una parte, la Grecia mítica y clásica; en segundo término, el Renacimiento y la Modernidad ilustrada prolongada incluso hasta el mismo corazón del Positivismo decimonónico; y en tercer lugar, el pensamiento contemporáneo, en el que ha conocido un averdadera floración el pensamiento dirigido a tomar en peso las realidades que representan la ciencia, la técnica, la tecnocracia y, últimamente, el desarrollo de todio lo que cabe englobar por referencia a la *Cíber Sociedad* y a la *Inteligencia Artificial*. En lo que resta de este epígrafe, daremos unas referencias básicas respecto de cada uno de los hitos a los que se acaba de hacer mención.

Una primera cala interesante en el proceso reflexivo acerca de la técnica nos habría de emplazar en la tradición cultural y filosófica griega. Ante todo, habríamos de arrancar con el abordaje mítico y poético que lleva a efecto Hesíodo en sus obras (Hesíodo 1982), donde nos ofrece el mito de Prometeo —«el que prevé, el que ve de antemano»—, que habrá de servir de paradigma del quehacer técnico, cuando menos, en el imaginario colectivo de Occidente.

Aunque, de hecho, Prometeo era venerado en Atenas como patrón de la industria, de la cerámica y de la artesanía, su figura resulta ambivalente, al igual que lo es la propia técnica, como indicaremos en un apartado posterior de este trabajo. Así, Prometeo, en cuanto trasunto de la capacidad técnica, aparece a primera vista como un indiscutible benefactor de la humanidad; a la que, de hecho, había regalado el fuego, epítome de la capacidad humana para producir, transformando en beneficio propio los recursos que la naturaleza le ofrece. Sin embargo, junto a este activo, al propio tiempo, Prometeo constituye también el ejemplo prototípico de quien actúa movido desde la *ubrys* y la desmesura, al haber tenido la osadía de desobedecer los mandatos de los dioses. Por consiguiente, tiene también como *debe*, el hecho de haber sido el responsable de muchos de los males que desde entonces vienen afligiendo a la humanidad. Y de manera muy significativa, siempre que ésta se conduce de manera imprudente y actúa de forma irresponsable con la técnica.

En efecto, Prometeo había hecho un impagable regalo a los hombres. Y esto habrá que contabilizarlo de manera inequívoca en su *haber*. Sin embargo, sólo lo pudo llevar a término su proeza, una vez hubo robado el fuego sagrado para vengarse de Zeus. Y aqueí esta la contrapartida y la fuente de la ambivalencia que el desarrollo tecnológico siempre conlleva.

El castigo divino no se hizo esperar: enterado el padre de los dioses del robo llevado a efecto por el titán Prometeo, le aplicó un doloroso suplicio, al atarlo a una columna y dejarlo expuesto a la voracidad de un águila que se aplica con

insistencia insufrible a devorarle el hígado que tiene la capacidad de regenerarse y de quedar siempre expuesto y a merced del águila por toda la eternidad. Los hombres, por su parte, hubieron de recibir también su correspondiente cuota de castigo, a resultas del pecado prometeico: de una parte, Zeus los sanciona con la creación de la mujer —ilustrada en el mito de Pandora—; de otra, sobre aquéllos planeará siempre lo que representa la figura del hermano de Prometeo, Epimeteo —«el que tarda en pensar, el que piensa tarde»—, con todo lo que conlleva de expresión de la tropeza humana.

En definitiva, el Prometeo que Hesíodo saca a escena por vez primera en su *Teogonía* y en su *Los trabajos y los días*, representa a la humanidad misma, que con el desarrollo tecnológico puede, ciertamente, someter a su desiognio la naturaleza. Ahora bien —y aquí está la gran metáfora que ha venido requiriendo, y aún lo habrá de demandara en el futuro, de la adecuada hermenéutica interpretativa de acuerdo a los distintos contextos a que su desarrollo vaya dando lugar—, si aquel saber resultare desmesurado, habrá de redundar, a la postre, en mayores problemas y desgracias de los que quisiera resolver. Y ello, al margen de la buena voluntad que se haya de atribuir; e incluso asumiendo la más recta de las intenciones que imaginar se pueda como motivadora del desarrollo tecnológico en cada momento histórico que se haya de considerar.

Con todo, prolongando la potente intuición hesiódica que acabamos de presentar, ahora ya en clave filosófica, cabría hacerse eco de lo que Platón —en *La República* (Platón 1988) y en el diálogo *Protágoras* (Platón 2017)— y Aristóteles —de una parte, en la *Ética a Nicómaco* (Aristóteles 1985); y de otra en la *Metafísica* (Aristóteles 2011)— aportan en lo tocante a la consideración del saber técnico y de la aplicación de la técnica a la producción transformadora de la naturaleza, como respuesta a necesidades y aspiraciones humanas. A tono con el alcance y el enfoque de este artículo, no cabe a este respecto más que animar a quien esté interesado en ello a que abunde por su cuenta en las fuentes señaladas, al objeto de perfilar algunos de los asuntos más sugerentes en la consideración filosófica de la técnica como saber y actividad poyético–productiva.

Tras haber hecho referencia al momento poético y reflexivo que la Grecia clásica supone con respecto a la técnica y a lo que con ella se relaciona, en segundo término, sería preciso parar mientes en la extensa etapa que, arrancando con el Renacimiento y transitando por la Modernidad ilustrada, se prolongada hasta los terrenos del Positivismo decimonónico. A lo largo de esos cuatro siglos, junto a la fascinación por la ciencia (Bacon 1985); al lado del

legítimo orgullo que brotaba de la constatación del indiscutible desarrollo económico y cultural que la *máquina* —convertida en gran metáfora de la Revolución Industrial, extensible a la consideración del hombre en su conjunto (La Mettrie 1987; López Corredoira 2019)—, trajo consigo, fruto del desarrollo técnico; y en sintonía con el optimismo epocal, mantenido incluso pese al hecho de tener que reconocer las insuficiencias de un desarrollo desigual y, hoy en día, problemático en su sostenibilidad; junto a todo ello, empieza a abrirse paso con fuerza algo que, como ya hemos indicado, había conocido una primera formulación en el pensamiento griego antiguo, pero que ahora se iba a plantear con mayor criticismo. A saber: de una parte, el sentido de lo que la técnica puede acabar trayendo como consecuencia, cuando aquélla se aleja de la racionalidad ética (Rousseau 1962; Shelley 2002). Y de otra, la recomendación respecto a que los científicos debieran asumir la responsabilidad derivada de sus creaciones.

El tercer momento, como ya indicábamos, cabe situarlo, precisamente, en la contemporaneidad. Es ahora cuando de manera más abierta y sistemática ha venido recibiendo la técnica una cumplida consideración filosófica. Destacan entre los pensadores contemporáneos, entre nosotros, José Ortega y Gasset (2014); y alineados en la estela reflexiva que aborda la técnica como práctica social y el impacto que ésta haya de tener en la dinámica vital de lo humano, cabe señalar a Martin Heidegger (2017), el ya citado Jürgen Habermas (1985), Hans Jonas (1995), Jacques Ellull (1960); así como a algunos de los actuales voceros del transhumanismo y del post humanismo (Mahon 2017) a los que habremos de referirnos más adelante.

Con todo, insistamos, el tenor de nuestro estudio, si bien ha de ser enmarcado en el contexto al que hemos querido hacer referencia en este epígrafe, se orienta a dilucidar si la racionalidad ética puede aportar claves (Coeckelbergh 2021), en forma de un denominador común, capaz de contribuir a encauzar a favor de lo humano el desarrollo tecnológico propio de la *Digitalización*. De lo que se trata, en definitiva, no es sino de apostar por una suerte de *Humanismo Digital* que ubique a la persona humana en el centro de todo el proceso que estamos viviendo con el advenimiento de la *Cíber Sociedad*, en el marco de la *Cuarta Revolución Industrial* (Schwab 2017).

Cuando, como diremos inmediatamente, el ser humano corre el peligro de verse amenazado en toda regla desde planteamientos más o menos delirantes (Cortina y Serra 2016), conviene insistir en la dignidad que le cabe atribuir, a tenor de una idiosincrasia y una peculiaridad que, por una parte, lo abre a la *Axiología,* le permite adentrarse en el mundo de los valores y tratar de vivirlos

(Martínez Díez 2021); por otro lado, lo capacita para apuntar hacia ámbitos de trascendencia, situados más allá de sí mismo y de lo dado, tanto en el ámbito de lo material cuanto en el de lo artificialmente producido o susceptible de acabar siendo producido o capaz de emerger en el futuro, más o menos *sua sponte*, como indicaremos a continuación, al referirnos a la dinámica del aprendizaje no supervisado propia de los algoritmos presentes en el fenómeno del *Machine Learning*.

Antes de ello, sin embargo, debemos dar una breve descripción de cuáles han sido las condiciones que posibilitaron el advenimiento de un contexto tan ambivalente —por intrigante y retador al propio tiempo— como es el que nos está tocando vivir en este momento histórico.

§3. Condiciones técnicas que posibilitaron la *Digitalización*

En este primer cuarto del siglo XXI, la humanidad está viviendo una circunstancia compleja: a la vez, convulsa y fascinante. De una parte, se ve enfrentada al reto de consolidar dinámicas de sostenibilidad en los tres ámbitos a través de los que se encauza la dimensión social del ser humano sobre la Tierra: la estabilidad necesaria para desplegar unos procesos político–culturales justos y respetuosos con la dignidad de las personas; un desarrollo económico equitativo que favorezca el florecimiento de los individuos y el progreso de los pueblos; y la preservación del sistema ecológico de la biosfera que, por lo demás, constituye la condición de posibilidad de todo lo anterior.

Por otra parte, se constata con orgullo un espectacular avance en la dimensión técnica de la vida humana en sociedad, cuyo epítome tal vez hayamos de colocarlo en el desarrollo que la *Inteligencia Artificial* ha venido conociendo a lo largo de los últimos setenta y cinco años, a través de las diferentes etapas en que se periodifica su línea de avance y tendencia de futuro: *Artificial Narrow Intelligence* —ANI—, *Artificial General Intelligence* —AGI— y *Artificial Super Intelligence* —ADI— (Kaplan y Haenlein 2020). Pues, en efecto, dejando al lado aproximaciones propias de la literatura fantástica, la ciencia ficción y el entretenimiento que Hollywood ofrece, lo cierto es que la historia del proceso que ha llevado desde la resolución de *Enigma* por parte de Alan Turing hasta la actualidad, es espectacular (Haenlein y Kaplan 2019).

Una primera mirada de cautela debiera llevarnos a preguntar por el sentido, la dirección y, sobre todo, la finalidad al servicio de la cual se estructura la digitalización de la economía en general, y la innovación y el desarrollo tecnológico en concreto (Brusoni y Vaccaro 2017). Ciertamente, compartimos la afirmación de Martin, Shilton y Smith (2019), cuando afirman que «neither

the efficiencies produced by the use of digital technology, nor enhanced financial return to equity investors solely justify the development, use, or commercialization of a technology» (Martin, Shilton y Smith 2019, p. 309). Por consiguiente, si las razones económicas y financieras no son suficientes para legitimar el desarrollo tecnológico como meta deseable hacia la que tender, habremos que buscar aquella legitimación moral en la mejora de las condiciones de vida y en la búsqueda del bienestar y del progreso de la gente.

En todo caso, el desarrollo tecnológico y la digitalización es un hecho. Por lo demás, las condiciones técnicas que posibilitaron el desarrollo de la *Inteligencia Artificial* resultan fáciles de identificar. De una parte, hay que hacer referencia al aumento del poder de computación; de otra, al incremento de la capacidad de almacenamiento; en tercer lugar, a la proliferación de datos —los famosos *Big Data* o macro datos—; y sobre todo, los avances en el desarrollo de los algoritmos que, entre otras cosas, han favorecido el advenimiento de la tecnología cognitiva y el *Machine Learning* (Alpaydin 2020), capaz de detectar patrones y aprender a hacer predicciones y recomendaciones mediante el procesamiento de los datos, sin tener que recibir órdenes explícitas del exterior.

Como señala, con agudeza Martin: «Algorithms silently structure our lives. Algorithms can determine whether someone is hired, promoted, offered a loan, or provided housing as well as determine which political ads and news articles consumers see» (Martin 2019, p. 835). Tienen, por consiguiente, una relevancia tan decisiva en tantas esferas de la vida del individuo en sociedad que se hace necesario reflexionar con sistema acerca de las conexiones éticas de su diseño, por parte de los ingenieros informáticos que los imaginan y programan, así como respecto de la responsabilidad de las empresas que los desarrollan e implementan y de los managers que las dirigen (Kahlil 1993).

Porque, en efecto, los algoritmos del *Machine Learning* se van adaptando a los nuevos datos y experiencias, para mejorar la eficacia de los análisis —tanto predictivos como prescriptivos— que llevan a efecto. Y naturalmente, en el proceso y en las posteriores aplicaciones que se hayan de hacer, pueden emerger dilemas éticos, al entrar en juego valores y principios morales, e incluso derechos de la persona que podrían verse conculcados (Mittelstadt et al. 2016).

En el proceso del *Machine Learning* se va generando un aprendizaje que, a su vez, puede ser de tres tipos: *Supervised Learning, Unsupervised Learning* y *Reinforced Learning*. En el primer tipo de aprendizaje, el algoritmo usa datos y la retroalimentación aportada por los humanos para aprender cuál es la relación

que hay entre un determinado input y un correlativo output. Este tipo de aprendizaje se suele utilizar cuando se conoce la clasificación de los datos y el tipo de conducta que se quiere predecir, pero se quiere que sea el algoritmo quien calcule el nuevo dato —por ejemplo: mes del año, tipo de interés y precio de la vivienda—. En el *Unsupervised Learning*, un algoritmo explora datos input sin que se le indique una concreta variable output. Por ejemplo, explora la demografía de los clientes para identificar patrones. Se utiliza cuando no se sabe cómo clasificar los datos y se quiere que sea el algoritmo quien lo haga. Por su parte, cuando un algoritmo aprende a desarrollar tareas, tratando de maximizar las recompensas que recibe en función de lo que consigue hacer —por ejemplo, recibe puntos cuando aumenta el retorno de una cartera de inversión—, estaríamos en presencia del *Reinforced Learning*, propiamente dicho.

Por su parte, el denominado *Deep Learning* es un tipo especial de *Machine Learning*, capaz de procesar un rango más amplio de recursos de datos y que, aunque requiere menor concurso humano, ofrece resultados más precisos que los métodos y sistemas anteriores (LeCun, Bengioi y Hinton 2015). El *Deep Learning* interconecta, en una especie de red neuronal, distintos niveles de software, conocidos precisamente como «neuronas». Los principales modelos de *Deep Learning* son, de una parte, el *Convolutional Neural Network* —CNN—; y de otra, la *Recurrent Neural Network* —RNN—. La *Convulational Neural Netwok* (CNN) recibe una imagen —por ejemplo, la letra «A»— y la procesa como una colección de *pixels*. En los niveles profundos identifica rasgos únicos, por ejemplo, las líneas que configuran la «A» y el sistema aprende a clasificar otras letras «A». La CNN sirve, por ejemplo, para diagnósticos médicos… También se puede utilizar para detectar el logo de una empresa en los medios, permitiendo con ello afinar campañas de marketing… o para identifica productos defectuosos en la línea de producción… Por su parte, la RNN, que es capaz, por ejemplo, de predecir la probabilidad de que aparezca una determinada palabra en una frase, se utiliza, entre otras cosas, para detectar transacciones fraudulentas en tarjetas de crédito.

Este escenario tecnológico en el que nos encontramos —con una *Inteligencia Artificial* capaz de desplegar el *Deep Learning*—, y desde el que habremos de tratar de dar salida a los impresionantes desafíos a los que hemos hecho referencia más arriba, es ambivalente (Johnson 2015). Es capaz de hacer que la humanización de la vida progrese; pero también está siempre presente el peligro de que —en línea con la acertada intuición que habla de un *responsibility gap* (Matthias 2004)— no todo redunde en beneficio de las personas y de la sociedad en su conjunto. Más allá de lo que De George denominara el mito de

la *amoral computing and information technology* —ACIT— (De_George 2003), la técnica tiene una indiscutible *dimensión moral*, por más que haya quienes quieran mantener aún la tesis separatista entre Ética y Técnica —al modo en que, en otro contexto, años atrás se utilizaba el mismo esquema interpretativo respecto a la Economía y la Ética— (Martin y Freeman 2004); y la tecnocracia plantea en sí misma serios interrogantes, que se sustancian en debates muy vivos acerca de cómo conjurar los riesgos y las amenazas implícitas en el desarrollo tecnológico (Kaplan y Haenlein 2020).

§4. La posibilidad, no tan lejana, de una distopía cibernética

Señalemos algunas de las citadas amenazas, bien que sin ánimo alguno de exhaustividad (Coeckelbergh 2021). En primer lugar estaría la más grave, a la que ya hemos aludido *supra* y que, ciertamente, constituye una pretensión formidable: la denominada *revolución transhumanista*, con la que se estaría apostando expresamente por una supuesta *mejora* de la raza humana —*Human Enhancement*—, (Cortina y Serra 2015; Vilaplana Guerrero 2019) cuando no por la creación de una especie nueva (Bostrom y Savulescu 2009; Ferry 2016; Mahon 2017; Lumbreras 2020). Las posibilidades técnicas parecen estar al alcance de la mano, mediante la convergencia de las conocidas como NBIC, esto es: la Nanotecnología, la Biotecnología —*Crispr–Cas9*—, la Tecnología de la Información y las Ciencias Cognitivas.

Ahora bien, son tantas y de tal calado las derivadas éticas que emergen ante el planteamiento de esta posibilidad, que, para enfrentarse a sus previsibles riesgos y peligros potenciales con un buen antídoto, se debería, de una parte, estimular en la ciudadanía global el despliegue y el desarrollo de capacidad de pensar críticamente (Brookfield 1987; Inch y Warnick 2001) sobre estos asuntos. Para ello, como primera providencia, se habría de apostar por una educación rigurosa que ayudara a un discernimiento maduro en ámbitos tan complejos y de tanta trascendencia (Radermacher 2018). Por otro lado, sería de desear que se pudiera llevar a efecto un amplio debate, convenientemente articulado y alejado, en todo caso, tanto de tesis apocalípticas respecto del negro futuro que a la humanidad le cabría esperar, si nos atenemos a las previsiones más agoreras e infundadas; cuanto del simplismo propio de planteamientos, ingenuos en exceso, que desistiendo incluso de identificar la dimensión ética del hecho tecnológico, tienden a reificarlo e, incluso, a inmunizarlo frente a la crítica menos acre.

En todo caso, en el proceso de diálogo al que aludimos como antídoto frente a posturas irresponsables en referencia a las consecuencias del

dersarrollo tecnológico en el marco de la *Cíber Sociedad*, habrían de poder participar, no sólo científicos y tecnólogos, sino también, juristas, políticos y otros representantes de cuerpos y asociaciones del mundo profesional e incluso de ciudadanos particulares. En dicho diálogo deberían intervenir no sólo las empresas, los *lobbies* y los representantes del poder económico, sino también diversos agentes de la *sociedad civil*.

Naturalmente, en dicho proceso dialógico se habría de dar cabida como elemento imprescindible a *la voz de la Ética* como Filosofía Moral; y a las distintas tradiciones culturales y religiosas de la Humanidad (Buchanan 2011). A este respecto, la aportación de instituciones como Globethics —a la que venimos haciendo referencia implícita mediante una profusa citación de referencias, pulicadas desde ella— y de otros *Think–Tanks* o centros de investigación similares, resultará de especial relevancia.

Porque, como decimos, es tanto lo que está en juego —lo más obvio, la dignidad de las personas, de un lado; y, de otro, la exacerbación de diferencias injustas entre personas, pueblos y culturas—, que la amenaza de una distopía de tal magnitud —ya sea en clave *huxleyana* (Huxley 2011), ya *orwelliana* (Orwell 2016)— debiera, cuando menos, hacernos formular algunos principios morales o, cuando menos, ciertos criterios de actuación práctica como los que Isaac Asimov, con una capacidad de intuición visionaria insólita en 1950, año de la publicación de su famoso *I, robot*, estampaba en el frontispicio de dicha novela de ciencia–ficción (Asimov 2020), dando referencia de un supuesto Manual de Robótica, en su quincuagésimosexta edición, del año 2058, bajo el rótulo de: «Las Tres Leyes de la Robótica». A saber: «1. Un 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» (Asimov 2020: 7).

De una manera más general, y en otro contexto, me hube de referir a dos máximas morales de puro sentido común que merece la pena rescatar. A saber: que *no todo lo éticamente deseable es técnicamente posible* en un momento histórico determinado; pero, al propio tiempo, que *no todo lo técnicamente posible resulta siempre éticamente deseable* (Fernández Fernández 1994).

En todo caso, sin tener por qué llegar a extremos distópicos tan peligrosos como los que cabe intuir en la apuesta por la *Singularity,* por *la Mort de la mort* (Alexandre, 2011) y por lo que el *Transhumanismo* representa (Buchanan 2011; Mahon 2017; Beorlegui 2018; Arana 2019; Lumbreras 2020; Martínez Díez

2021), es posible encontrar otros elemenstos que, sin duda, plantean también consideraciones inquietantes desde el punto de vista moral. Así, por ejemplo, junto a la ya aludida *Cybersecurity* (Hurlburt 2018), estaría el aumento de la huella energética, derivado de las necesidades que las nuevas tecnologías y la cultura digital trae consigo; la posible sustitución de procesos democráticos por lo que se ha dado en llamar el *technological solutionism*; la pérdida de confianza social, con el aumento de la polarización y el fanatismo que ello podría representar. Ello podría ser debido, entre otras cosas, a la proliferación de los bulos y *fake news*; así como a la denominada *economía de la atención*, en virtud de la cual, las personas están siendo constantemente bombardeadas, no sólo con informaciones de tipo comercial, más o menos subliminales, sino también con otro tipo de mensajes y de manipulaciones contrarias a una convivencia pacífica.

Además de lo que va indicado, entre los desafíos que este *Cyber Space* nos plantea, cabe indicar, la posibilidad de terminar viviendo en una sociedad excesivamente controlada y controladora en exceso; el impacto de la robotización en las cadenas productivas, con la aparición de un nuevo taylorismo digital y la redundancia de muchos empleos, con el impacto social y personal que ello habrá de suponer; la autonomización creciente de las máquinas y la correlativa dilución de la responsabilidad por parte de los sujetos humanos; una creciente concentración de poder en pocas manos, por lo demás, envuelta en un contexto de creciente opacidad; posibilidad de múltiples malas prácticas e incluso de crímenes cibernéticos de nuevo cuño; pérdida de privacidad; la utilización injusta de algoritmos que, en sus tomas de decisiones contribuyan, más o menos expresamente, a ocultar, dar legitimidad o a perpetuar sesgos injustos y procesos de discriminación inaceptables.

Por fortuna, no partimos de cero, sino que hay ya una abundante panoplia de iniciativas orientadas a tratar de poner coto a aquellos escenarios y a hacer frente a los desafíos morales que de ellos se desprenden.

§5. Algunas propuestas en materia de *Cíber Ética*

Conviene resaltar el hecho de que, precisamente como respuesta a unas amenazas tan serias como las que acabamos de enumerar, es unánimemente reconocida la necesidad de tomar en consideración, no sólo la *dimensión ética* del proceso por el que va evolucionando y configurándose esta especie de sociedad cibernética, sino también la perspectiva propia de lo que cabría denominar como *Cíber Derecho* (Duggal 2018; Toolen 2018).

No se trata, a buen seguro, ni de descubrir nuevos valores nunca antes

conocidos; ni de encarecer la deseabilidad de virtudes inéditas. Por supuesto, tampoco habrán de tener por qué hacer acto de presencia normas y reglas de nuevo cuño para ser aplicadas en el concierto social planetario al que nos abocan las nuevas realidades. Más bien, lo que procedería sería articular un acervo moral bien conocido —responsabilidad, libertad, justicia, paz, seguridad, igualdad, participación, transparencia, respeto, etc.— (Stückelberger, Fust y Obiora 2016); de manera creativa, relacional y, sobre todo, ajustada a las circunstancias que la *Cíber Sociedad* y el desarrollo de la *Inteligencia Artificial* representan.

De hecho, no son pocas las iniciativas en marcha con referencia a la propuesta de guías éticas para encauzar desde las mejores prácticas técnicas y atender a los requerimientos éticos de la *Inteligencia Artificial* desde criterios, principios y valores morales. En los últimos cinco años se ha ido desplegando una amplia floración de Guías y documentos (Jobin, Ienca y Vayena 2019; Larsson 2020), tanto desde compañías privadas que dan cuenta de buenas prácticas (Wang, Xiong y Olya 2020), cuanto desde el punto de vista de las propuestas que llevan a efecto profesionales de la informática, desarrolladores de sistemas, o empresas de software; y por supuesto desde los planteamientos que se vienen realizando a nivel administrativo (Cerillo i Martínez 2019) y político, que buscan situar la tecnología, la *Inteligencia Artificial* y los algoritmos al servicio de las personas y de los valores humanos.

A modo de ejemplo ilustrativo, junto a documentos de alto calado político, como los producidos desde la Unión Europea (European Commission 2018; European Commission 2020), podemos hacernos eco también de otros dos recientes informes, en los que, en línea con lo que va dicho, se insiste en la *dimensión ética* de la tecnología y la digitalización en esta sociedad digitalizada. De una parte, el informe ENIA del gobierno de España, donde se aborda la *Estrategia Nacional de Inteligencia Artificial. España Digital 2025*, y en el que el eje estratégico 6 se dedica explícitamente a: «Establecer un marco ético y normativo que refuerce la protección de los derechos individuales y colectivos, a efectos de garantizar la inclusión y el bienestar social» (Gobierno de España 2020, pp. 64–70). Por otra parte, un reciente informe de la OECD sobre la *smart–mobility*, tomando distancia de cualquier planteamiento tecno–centrista, aboga expresamente en el propio título por la construcción de unas *human–centric smart–cities* (OECD 2020).

Esta apuesta por la *Cíber Ética* ha tenido también su traducción en el ámbito académico. Denomínese *Ética Tecnológica* (Martin, Shilton y Smith 2019), *Ética Digital, Data Ethics* (Floridi y Taddeo 2016), *Ética de la Inteligencia Artificial*

(Kaplan y Haenlein 2020; Baker–Brunnbauer 2020; Coeckelbergh 2021), *Ética de los Algoritmos* (Mittelstadt *et al.* 2016; Monasterio Astobiza 2017; Martin 2019)… o con cualquier otra variación sobre el tema, la academia avanza en la necesaria reflexión y en las propuestas respecto a cómo aprovechar las circunstancias y las posibilidades que la digitalización está poniendo en manos de la humanidad en el día de hoy para construir un mundo más justo, más sostenible y, sobre todo, más plenamente humano para todos; donde cada uno pudiera florecer y desarrollarse como persona.

En un reciente trabajo, Mick Ashby llega a afirmar: «We are the only generation that has the chance to steer the fate of future generations of humanity towards being collectively ruled, potentially for eternity, by benevolent super–ethical systems that create a stable cyberanthropic utopia for us, effectively and ethically minimizing human suffering and environmental problems. The alternative is to allow hubris, insatiable greed, and super–unethical systems to extinguish our rights and freedoms, and either enslave most of us in a cybermisanthropic dystopia or cause the extinction of our species to become a footnote in Gaia's geologicval record» (Ashby, 2020, pg. 325).

Desde este planteamiento optimista, Ashby ofrece una propuesta Ética aplicada al diseño e implantación de lo que se denomina *Super–Ethical Systems*, a partir del conocido como *Good Regulator Theorem*. Mick Ashby (2020) llega a decir, no sólo que «the implementation of super–ethical systems is identified as an urgent impetrative for humanity to avoid the danger that superintelligent machines might lead to a technological dystopia» (Ashby 2020, p. 1), sino que, partiendo del *Good Regulator Theorem*, si se atiende a los nueve requisitos que él presenta en su abordaje del problema —Propósito, Verdad, Variedad de acciones, Predecibilidad, Inteligencia, Influencia en el sistema, Ética y reglas prioritarias, Integridad de todos los subsistemas y Transparencia—, se estaría en condiciones de poner en funcionamiento lo que él denomina un regulador ético y eficiente.

§6. Hacia un factor común para la *Cíber Ética* y la *Ética de la Inteligencia Artificial*

En el ya citado trabajo del año 2019, Jobin, Ienca y Vayena estudian un abundante *corpus* de 84 guías éticas sobre Inteligencia Artificial, tratando de encontrar una especie de factor ético común. Tras llevar a efecto el análisis de los documentos, de codificar sus contenidos y de cuantificar el número de veces que aparecen referidos, ofrecen el siguiente listado de principios ético, por

orden de prelación, en función del número de documentos en los que se mencionan:

1. *Transparencia* y términos conexos —transparencia, explicabilidad, intelegibilidad, interpretabilidad, comunication, divulgación y presentación d ela información—, aparecen en 73 de los 84 documentos analizados;
2. *Justicia and Equidad* —justicia, equidad, consistencia, inclusión, igualdad, (no–) sesgos, (no–) discriminación, diversidad, pluralidad, accesibilidad, reversibilidad, remedio, reparación, acceso y distribución—, aparecen en 68 de entre 84;
3. *No–maleficencia* —no maleficencia, securidad, daño, protección, precauciónn, prevención, integritidad (física o mental), no–subversión—, se reflejan en 60 de entre 84 guías; *Responsabilidad* —responsabilidad, confianza, actuar con integridad—, también, se reflejan en 60 de los 84 documentos analizados;
4. *Privacidad* y sus términos conexos —privacidad, información personal o privada—, aparece en 47 de 84; *Beneficencia* —beneficio, beneficencia, bienestar, paz, bien social, bien común— se menciona en 41 de 84 guías;
5. *Libertad y Autonomía* —libertad, autonomía, consentimiento, elección, autodeterminación, empoderamiento—, en 34 de 84;
6. *Veracidad* en 28;
7. *Sostenibilidad* —sostenibilidad, medioambiente (natural), energía,)recursos energéticos— en 14;
8. *Dignidad* en 13; y
9. *Solidaridad* —solidaridad, seguridad social, cohesión— en 6 de entre las 84 (Jobin, Ienca y Vayena 2019, p. 7).

Por su parte, el Grupo de Alto Nivel de la Comisión Europea ofrece cuatro principios éticos, como requisitos para una IA digna de confianza: a) Respeto a la autonomía humana; b) Prevención del daño; c) Equidad; y d) Explicabilidad. Dichos elementos se ven complementados por referencia a los otros siete presupuestos siguientes: 1) Agencia y supervisión humana; 2) Solidez técnica y seguridad; 3) Respecto a la privacidad y adecuada gestión de los datos; 4) Transparencia; 5) Diversidad, no discriminación y equidad; 6) Bienestar social y

medioambiental; y 7) Responsabilidad.

Como se observa, los principios morales, los valores éticos, los criterios que habrían de inspirar las leyes y reglamentaciones que habrían de ser promulgadas (Duggal 2018), incluso, las virtudes (Stückelberger 2018) y las prácticas que debieran desarrollarse para una buena interacción personal en el marco de la la sociedad digital son aspectos en los que se produce una muy significativa convergencia teórica.

Con todo, queda un camino interesante que seguir recorriendo: de un lado, es evidente un desajuste entre los avances tecnológicos y la codificación legislativa; de otra parte, parece necesario seguir avanzando en estudios y providencias que ayuden a pasar de los principios a los procesos en la gestión de la digitalización, la *Cíber Sociedad* y la *Inteligencia Artificial*. Finalmente, como señala Larsson (2020), habrá que insistir en la necesidad de abundar en el estudio de las posibles nuevas aplicaciones de la ciencia de los datos que la *Inteligencia Artificial* despliega cada día.

§7. Conclusión: Una apuesta axiomática por el *Humanismo Digital*

El desarrollo tecnológico y la *Inteligencia Artificial* han propiciado la emergencia del entorno de lo que se denomina *Cíber Sociedad* o *Sociedad Digital*. Sin embargo, las expectativas de futuro que se abren ante la humanidad resultan ser ambivalentes: si, de una parte, son fascinantes —con toda la carga de emoción positiva que el término tiene en su raíz etimológica—; de otra, no pueden sino aparecer como *formidables* —con la inevitable carga de negatividad que la referencia etimológica trae consigo en este caso: de *formido*, que, en latín, quiere decir *miedo*.

Tomar las riendas de la *digitalización* y hacer que la *Inteligencia Artificial* se ponga a favor de las personas y de la humanidad en su conjunto, en una suerte de *Humanismo Digital*, está en nuestras manos. Y la dimensión ética del desarrollo tecnológico aparece como uno de los desafíos más imperiosos —tal vez a la par con el del problema ecológico— de cara a los próximos años.

Los principios morales tradicionalmente asociados a la *Bioética* parece ser que podrían encontrar buen acomodo en el nuevo escenario. *No maleficencia, Beneficencia, Autonomía* y *Justicia*, en definitiva, siguen siendo expectativas morales innegociables. Tal vez, como hemos visto, se deban complementar con otros dos criterios éticos fundamentales: *Transparencia y Explicabilidad*, de una parte; y *Responsabilidad*, de otra. Ello sea dicho, sin perjuicio de que, en este

particular, quepa también ensayar abordajes complementarios —nunca sustitutivos— al que va dicho. Tal es, en nuestro caso, uno en el que venimos trabajando desde hace tiempo, alineado con la *Ética del Cuidado* (Villegas Galaviz y Fernández Fernández 2021, en prensa).

Con todo, lo cierto es que conviene seguir dándole vueltas a la reflexión ética, desde cualquiera de sus enfoques y paradigmas, porque, asumiendo que los algoritmos que animan los sistemas donde la *Inteligencia Artificial* se aplica, nunca son neutros —y, sin duda, aspirando a que se evite todo tipo de atropellos contra la dignidad personal o el respeto debido a tradiciones culturales legítimas—; la asepsia moral tampoco es el valor supremo al que cupiera apelar. Más bien ocurre todo lo contrario: si la gente actúa de ordinario *sub specie boni*, es precisamente la reflexión moral y el discernimiento ético los que hayan de discriminar de manera fundada y razonable, entre lo bueno, lo malo y lo mejor…

En consecuencia, tal vez merezca la pena mantener la diferencia entre la *Inteligencia Artificial* —mucho más potente, capaz de almacenar datos y de llevar a efecto operaciones y cómputos imposibles para cualquier persona— y la *Inteligencia Natural*. La *Inteligencia Artificial* es, en definitiva, un producto humano que supera a su productor… pero sólo en un aspecto de la ecuación… Es más *inteligente*, sin duda alguna; si por inteligencia se entiende lo que va dicho. Pero, en cambio, tal vez nunca pueda resultar ser más *lista*. Si con esta categoría queremos referirnos a algo que no es sino patrimonio exclusivo de nuestra *inteligencia sentiente*. Es imperfecta, falible, limitada y frágil; pero, al mismo tiempo, resulta ser también empática, emocional, poética, libre, responsable y abierta al Espíritu y a la Transcendencia.

En este punto cabe esperar que siempre haya de ser posible establecer una distinción cualitativa que marque la diferencia entro lo humano y lo no humano: ya sea esto meramente instrumental, ya nos hayamos de referir a proyectos transhumanistas o, directamente a planteamientos que busquen optar de manera abierta por lo post humano.

Mantener la bandera del humanismo en estos momentos pudiera incluso resultar un empeño, si no heroico, sí cuando menos, un tanto a contracorriente y discordante respecto al discurso que mayor eco parece encontrar en el gran relato que se va construyendo y en el que, entre otros protagonistas, destacan la *Digitalización*, la *Cíber Sociedad* y la *Inteligencia Artificial*. Con todo, merece la pena innovar desde la adecuada interacción hombre–máquina, hacia un nuevo *Humanismo Digital*. Al menos para quienes acepten la ardua tarea de perfeccionar de veras lo humano, sin traicionarlo.

REFERENCIAS

ALEXANDRE, Laurent (2011). *La Mort de la mort. Comment la technomedécine va bouleverser l'humanité.* J. C. Lattès.

ALPAYDIN, Ethem (2020). *Introduction to Machine Learning.* Cambridge (Massachusetts): The MIT Press.

ARANA, Juan (2019). «El futuro del hombre. ¿Contienen las propuestas del transhumanismo una respuesta satisfactoria?». En: *Inteligencia Artificial y Antropología Filosófica. ¿Es posible transferir la mente humana a un soporte no biológico?,* editado por Teodoro Sánchez–Avila Sánchez–Migallón. *Naturaleza y Libertad. Revista de estudios interdisciplinares* 12 (volumen monográfico): pp. 45–66.

ARISTÓTELES (1985). *Ética a Nicómaco.* Madrid: Centro de Estudios Constitucionales.

ARISTÓTELES (2011). «Metafísica». En Aristóteles, *Protréptico. Metafísica* (pp. 67–468). Madrid: Gredos.

ASHBY, Mick (2020). «Ethical Regulators and Super–Ethical Systems». *Systems* 8, no. 53: pp. 1–35.

ASIMOV, Issac (2020). *Yo, robot.* Barcelona: Edhasa.

BACON, Francis (1985). *Novum Organum: aforismos sobre la interpretación de la naturaleza y el reino del hombre.* Barcelona: Orbis.

BAKER–BRUNNBAUER, Josef (2021). «Management perspective of ethics in artificial intelligence». *AI and Ethics* 1: pp. 173–181.

BEORLEGUI, Carlos (2016). *Antropología Filosófica. Dimensiones de la realidad humana.* Madrid: Universidad Pontificia Comillas.

BEORLEGUI, Carlos (2018). *Humanos. Entre lo prehumano y lo pos– o transhumano.* Madrid: Sal Terrae y Universidad Pontificia Comillas.

BERGER, Peter L. y LUCKMANN, Thomas (1979). *La construcción social de la realidad.* Buenos Aires: Amorrortu.

BOSTROM, Nick, y SAVULESCU, Julian (2009). *Human Enhancement.* Oxford: Oxford University Press.

BROOKFIELD, Stephen D. (1987). *Developing critical thinkers: Challenging adults to explore alternative ways of thinking and acting.* San Francisco, CA: Jossey–Bass.

BRUSONI, Stefano y VACCARO, Antonino (2017). Ethics, Technology and Organizational Innovation. *Journal of Business Ethics* 143: pp. 223–226.

BUCHANAN, Allen (2011). *Better than Human. The Promise and Perils of Enhancing Ourselves.* Oxford: Oxford University Press.

CERILLO I MARTÍNEZ, Agustí (2019). «How can we open the black box of public administration? Transparency and accountability in the use of algorithms». *Revista Catalana de Dret Públic* 58: pp. 13–28.

COECKELBERGH, Mark (2021). *Ética de la Inteligencia Artificial.* Madrid: Cátedra.

CORTINA, Adela y SERRA, Miquel–Àngel (2015). *¿Humanos o posthumanos? Singularidad tecnológica y mejoramiento humano.* Barcelona: Fragmenta Editorial.

CORTINA, Adela y SERRA, Miquel–Àngel (2016). *Humanidad. Desafíos éticos de las tecnologías emergentes.* Madrid: Ediciones Internacionales Universitarias.

DE GEORGE, Rirchard T. (2003). *The ethics of information technology and business.* Malden: Blackwell Publishing.

DUGGAL, Pavan (2018). «Cyber Law and Cyber Ethics: How the Twins Need Each Other». En *Cyber Ethics 4.0. Serving Humanity with Values,* editado por Christoph Stückelberger y Pavan Duggal (pp. 55–68). Geneva: Globethics.net.

ELLUL, Jacques (1960). *El siglo XX y la técnica.* Barcelona: Labor.

ESQUIROL, Josep Maria (2011). *Los filósofos contemporáneos y la técnica. De Ortega a Sloterdiijk.* Barcelona: Gedisa.

EUROPEAN COMMISSION (20 de December de 2020). *WHITE PAPER. On Artificial Intelligence – A European approach to excellence and trust.* Obtenido de https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

EUROPEAN COMMISSION, y INDEPENDENT HIGH LEVEL EXPERT GROUP ON AI (20 de December de 2018). *Ethics Guidelines for Trustworthy AI.* Obtenido de https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

FERNÁNDEZ FERNÁNDEZ, José Luis (1994). «La Economía como oportunidad y reto de la Ética Profesional». En *Ética de las Profesiones,* editado por José Luis Fernández Fernández y Augusto Hortal Alonso (pp. 83–107). Madrid: Universidad Pontificia Comillas.

FERRY, Luc (2016). *La Révolution Transhumaniste. Comment la technomédecine et l'uberisation du monde vont bouleverser nos vies.* Paris: Plon.

FLORIDI, Luciano y TADDEO, Mariarosaria (2016). «What is Data Ethics?» *Phiosophical Transactions of the Royal Society A.* 374, no. 2083: pp. 1–5.

GOBIERNO DE ESPAÑA (Diciembre de 20 de 2020). *ENIA. Estrategia Nacional de Inteligencia Artificial. Versión 1.0.* Obtenido de Gobierno de España: https://portal.mineco.gob.es/RecursosNoticia/mineco/prensa/noticias/2020/201202_np_ENIAv.pdf

HABERMAS, Jürgen (1985). *Ciencia y técnica como ideología.* Madrid: Tecnos.

HABERMAS, Jürgen y HUSSERL, Edmund (1995). *Conocimiento e interés / La filosofía en la crisis de la humanidad europea.* Valencia: Unuiversitat de València.

HAENLEIN, Michael y KAPLAN, Andreas (2019). «A Brief History of Artificial Intelligence: On the Past, Present and Future of Artificial Intelligence». *California Management Review* 61, no. 4: pp. 5–14.

HEIDEGGER, Martin (2017). «La pregunta por la técnica». En *Filosofía, Ciencia y Técnica* (pp. 75–95). Santiago de Chile: Editorial Universitaria.

HEREDIA, Daniel (2018). «¿Qué constituye al ser humano como ser humano? Un intento de tomar parte en la búsqueda a una pregunta básica del quehacer filosófico». En: *Humanos e inhumanos. Qué nos asemeja y qué nos diferencia de las restantes especies animales,* editado por Teodoro Sánchez–Avila. *Naturaleza y Libertad. Revista de estudios interdisciplinares* 10 (volumen monográfico): pp. 143–154.

HESÍODO (1982). *Teogonía – Trabajos y Días – Escudo – Fragmentos – Certamen.* Madrid: Gredos.

HURLBURT, George (2018). «Toward Applied Cyberethics». *Computer* 51: pp. 80–84.

HUXLEY, Aldous (2011). *Un mundo feliz.* Barcelona: Penguin Random House.

INCH, Edward S., y WARNICK, Barbara H. (2001). *Critical Thinking: The use of reason in argument.* Boston: Allyn y Bacon.

JOBIN, Anna, IENCA, Marcello, y VAYENA, Effy (2019). «The global landscape of AI ethics guidelines». *Natura Machine Intelligence* 1, no. 9: pp. 389–399.

JOHNSON, Deborah G. (2015). «Technology with No Human Responsibility?». *Journal of Business Ethics* 127: pp. 707–715.

JONAS, Hans (1995). *El principio responsabilidad: ensayo de una ética para la civilización tecnológica.* Barcelona: Herder.

KAHLIL, Omar E. M. (1993). «Artificial Decision–Making and Artificial Ethics: A Management Concern». *Journal of Business Ethics* 12: pp. 313–321.

KAPLAN, Andreas y HAENLEIN, Michael (2020). «Rulers of the world unite! The challenges and oppotunities of artificial intelligence». *Business Horizons* 63: pp. 37–50.

LA METTRIE, Julien Offray (1987). *El hombre máquina.* Madrid: Alhambra.

LARSSON, Stefan (2020). «On the Governance of Artificial Intelligence through Ethics Guidelines». *Asian Journal of Law and Society* 7, no. 3: pp. 1–15.

LECUN, Yann, BENGIOI, Yoshua, y HINTON, Geoffrey (2015). «Deep Learning». *Nature* 521: pp. 436–444.

LÓPEZ CORREDOIRA, Martín (2019). Del hombre–máquina a la máquina–hombre. Materialismo, mecanicismo y transhumanismo. En: *Inteligencia Artificial y Antropología Filosófica. ¿Es posible transferir la mente humana a un soporte no biológico?,* editado por Teodoro Sánchez–Avila Sánchez–Migallón. *Naturaleza y Libertad. Revista de estudios interdisciplinares* 12 (volumen monográfico): pp. 179–190.

LÓPEZ MORATALLA, Natalia (2017). *Inteligencia Artificial. ¿Conciencia Artificial?* Madrid: Digital Reasons.

LUMBRERAS, Sara (2020). *Respuestas al transhumanismo. Cuerpo, autenticidad y sentido.* Madrid: Digital Reasons.

MAHON, Peter (2017). *Posthumanism: a guide for the perplexed.* New York: Bloomsbury Academic.

MARTIN, Kirsten (2019). «Ethical Implications and Accountability of Algorithms». *Journal of Business Ethics* 160: pp. 835–850.

MARTIN, Kirsten E., y FREEMAN, R. Edward (2004). «The Separation of Technology and Ethics in Business Ethics». *Journal of Business Ethics* 53: 353–364.

MARTIN, Kirten, SHILTON, Katie, y SMITH, Jeffery. (2019). «Business and the Ethical Implications of Technology: Introduction to the Symposium». *Journal of Business Ethics* 160: pp. 307–317.

MARTÍNEZ DÍEZ, Felicísmo (2021). *Humanos, sencillamente humanos. Desafíos del transhumanismo.* Madrid: San Pablo.

MATTHIAS, Andreas (2004). «The responsibility gap: Ascribing responsibility for the actions of learning automata». *Ethics and Information Technology* 6: pp. 175–183.

MITTELSTADT, Brent Daniel, ALLO, Patrick, TADDEO, Mariarosaria, WACHTER, Sandra, y FLORIDI, Luciano (2016). «The ethics of algorithms: Mapping the debate». *Big Data y Society* 3, no. 2: pp. 1–21.

MONASTERIO ASTOBIZA, Anibal (2017). «Ética algorítmica: Implicaciones éticas de una sociedad cada vez más gobernada por algoritmos». *Dilemata* 24: pp. 185–217.

OECD, y FORUM, I. T. (2020). *Leveraging digital technology and data for human–centric smart cities. The case of smart mobility. Report for the G20 Digital Economy Task Force.* OECD.

ORTEGA Y GASSET, José (2014). Meditación de la técnica. En J. Ortega y Gasset, *Ensimismamiento y alteración, Meditación de la técnica y otros ensayos* (págs. 55–138). Madrid: Alianza.

ORWELL, George (2016). *1984.* Barcelona: Penguin Random House.

PLATÓN (1988). *La República.* Madrid: Aguilar.

PLATÓN (2017). Protágoras. En Platón, *Diálogos* (pp. 235–300). Madrid: Gredos.

RADERMACHER, Ingo (2018). Cyber Ethics Requires Critical Thinking of Citizens. En *Cyber Ethics 4.0. Serving Humanity with Values,* editado por Christoph Stückelberger y Pavan Duggal (pp. 439–461). Geneva: Globethics.net.

ROUSSEAU, Jean Jacques (1962). *Discurso sobre las ciencias y las artes.* Madrid: Aguilar.

Schwab, Klaus (2017). *The Fourth Industrial Revolution.* New York: Crown Business.

SHELLEY, Mary W. (2002). *Frankenstein o el moderno Prometeo.* Madrid: Siruela.

STÜCKELBERGER, Christoph (2018). «Cyber Society: Core values and Virtues». En *Cyber Ethics 4.0. Serving Humanity with Values,* editado por Christoph Stückelberger y Pavan Duggal (pp. 23–53). Geneva: Globethics.net.

STÜCKELBERGER, Christoph, FUST, Walter, y OBIORA, Ike (2016). *Global Ethics for Leadership: Values and Virtues for Life.* Geneva: Globethics.net.

SWEENEY, Brian E. (2018). «The Nexus Between Cyber and Ethics». *National Defense. INDIA's Business & Technology Magazine* 103, no. 780: p. 38. https://www.nationaldefensemagazine.org/articles/2018/11/2/the-nexus-between-cyber-and-ethics

TOOLEN, Narayan (2018). «Law, Cyber Ethics and Technology». En *Cyber Ethics 4.0. Serving Humanity with Values,* editado por Christoph Stückelberger y Pavan Duggal (pp. 279–283). Geneva: Globethics.net.

VILAPLANA GUERRERO, José Domingo (2019). «Discusión crítica acerca de los principios que inspiran la supuesta necesidad y legitimidad del mejoramiento humano». En: *Inteligencia Artificial y Antropología Filosófica. ¿Es posible transferir la mente humana a un soporte no biológico?,* editado por Teodoro Sánchez–Avila Sánchez–Migallón. *Naturaleza y Libertad. Revista de estudios interdisciplinares* 12 (volumen monográfico): pp. 257–271.

VILLEGAS GALAVIZ, Carolina y FERNÁNDEZ FERNÁNDEZ, José Luis (2021). «Care Ethics in the era of Artificial Intelligence». En *Philosophy for Business Ethics*, editado por Guglielmo Faldetta, Edoardo Mollona, y Massimiliano M. Pellegrini. Palgrave (En prensa).

WANG, Yichuan, XIONG, Mengran, y OLYA, Hossein G. (2020). «Toward an Understanding of Responsible Artificial Intelligence Practices (HICSS 2020)». *Proceedings of the 53rd Hawaii International Conference on System Sciences* (pp. 4962–4971). Maui, Hawaii.

Towards *Digital Humanism* from a common denominator for *Cyber Ethics* and *Artificial Intelligence (AI) Ethics*

From an express option in favour of the human, the article is structured as a response to the following research question: Could we identify a common ethical denominator that could serve as a proposal for a Digital Humanism: for a situation in which humanity, freeing itself from avoidable suffering, manages to deploy its potential to achieve sustainable economic development and technical and political progress capable of giving rise to human flourishing? All of this, within the framework of the Cyber Society and placing the person and his or her dignity at the centre of the whole process related to the Artificial Intelligence of the present and the future.

On the basis of the Philosophy of Technology, and after taking into account the conditions of technical possibility of the Fourth Industrial Revolution, in an attempt to avoid the possibility of any dystopia, technically feasible but ethically vital, some proposals are considered in terms of Cyber Ethics and the Ethics of Artificial Intelligence; and some ethical principles are highlighted which seem to be serving as a common denominator from which to channel technological development in favour of humanity on a planetary scale.

It concludes by reiterating the commitment to Digital Humanism.

Keywords: Philosophy of Technique · Digitalisation · Cibernetic Dystopia · Ethical Principles.

Hacia el *Humanismo Digital* desde un denominador común para la *Cíber Ética* y la *Ética de la Inteligencia Artificial*

Desde una opción expresa a favor de lo humano, el artículo se estructura como respuesta a la siguiente pregunta de investigación: ¿Cabría identificar un común denominador ético que pudiera servir de propuesta para un Humanismo Digital: para una situación en que la humanidad, liberándose de sufrimientos evitables, lograra desplegar sus potencialidades para conseguir un desarrollo económico sostenible y un progreso técnico y político capaz de dar lugar al florecimiento humano? Todo ello, en el

marco de la Cíber Sociedad y poniendo a la persona y su dignidad en el centro de todo el proceso relacionado con la Inteligencia Artificial del presente y del futuro
Tomando base en la Filosofía de la Técnica y tras dar cuenta de las condiciones de posibilidad técnica de la Cuarta Revolución Industrial, tratando de conjurar la posibilidad de cualquier distopía, técnicamente factible, pero éticamente vitanda; se toman en consideración algunas propuestas en materia de Cíber Ética y de Ética de la Inteligencia Artificial; y se subrayan algunos principios éticos que parecen estar sirviendo de denominador común a partir del cual encauzar el desarrollo tecnológico a favor de la humanidad a escala planetaria.
Se concluye reiterando la apuesta por el Humanismo Digital.
Palabras Clave: Filosofía de la Técnica · Digitalización · Distopía cibernética · Principios Éticos.

JOSÉ LUIS FERNÁNDEZ FERNÁNDEZ es Catedrático en la Facultad de Ciencias Económicas y Empresariales (ICADE) y Director de la Cátedra de Ética Económica y Empresarial de la Universidad Pontificia Comillas, España. Doctor en Filosofía [≈ PhD] por la Universidad Pontificia Comillas. Fellow de la Caux Round Table, preside el Subcomité de Ética y Responsabilidad Social –CTN 165 SC2– de la UNE. Su investigación se central en la ética digital de negocios, en especial sobre la dimensión ética de Big Data, la analítica de la inteligencia artificial y la robótica. Ha escrito numerosos artículos y varios libros sobre Responsabilidad Social Corporativa, Gobierno Corporativo, Ética y Empresa.

INFORMACIÓN DE CONTACTO | CONTACT INFORMATION: Facultad de Ciencias Económicas y Empresariales (ICADE), Universidad Pontificia Comillas. C/ Alberto Aguilera, 23. 28015 Madrid, España. e–mail (✉): jlfernandez@icade.comillas.edu · iD: http://orcid.org/0000-0002-2344-7169.

HISTORIA DEL ARTÍCULO | ARTICLE HISTORY
Received: *21–May–2021*; Accepted: *29–June–2021*; Published Online: *30–June–2021*

COMO CITAR ESTE ARTÍCULO | HOW TO CITE THIS ARTICLE
Fernández Fernández, José Luis (2021). «Hacia el *Humanismo Digital* desde un denominador común para la *Cíber Ética* y la *Ética de la Inteligencia Artificial*». *Disputatio. Philosophical Research Bulletin* 10, no. 17: pp. 107–130.