

PARA UNA CONCEPCIÓN HUMANISTA DE LA TECNOLOGÍA

Reflexión crítica sobre la era bio-digital¹

Marta Bertolaso² y Alfredo Marcos³

Resumen

Los seres humanos somos seres técnicos. Nuestras vidas no se pueden lograr sin técnica. Al final de la modernidad, la simbiosis entre ciencia y técnica se ha vuelto ya tan estrecha que ahora podemos hablar de nuevas modalidades de lo técnico (tecnología, tecnociencia, biotecnología, antropotecnia), que nos otorgan más poder y responsabilidad que las simples técnicas tradicionales. Estas nuevas modalidades suscitan una reflexión filosófica que pone en juego nuestras concepciones ontológicas, epistemológicas, antropológicas y éticas. En este texto presentamos una reflexión crítica sobre las dos grandes propuestas antropotécnicas actuales, *bio* e *info*, y articulamos algunos argumentos y criterios para un adecuado humanismo tecnológico con énfasis en nuestra naturaleza relacional.

Palabras clave: era bio-digital, antropotecnias, humanismo tecnológico, dualismo, mejora humana, inteligencia artificial.

Abstract

Human beings are technical beings. Their lives cannot be accomplished without technique. At the end of modernity, the symbiosis between science and technique has become so tight that we can now speak of technology, and even technoscience, that gives us more power and responsibility than simple traditional techniques, as in the case of the new biotechnologies and anthropotechnics. In this text we present a critical reflection on the two major anthropotechnical proposals, that is, the bio-project and the info-project, and articulate some arguments and criteria for an adequate technological humanism with emphasis in our relational nature.

Keywords: bio-digital age, anthropotechnics, technological humanism, dualism, human enhancement, artificial intelligence.

1. Introducción

El ser humano es un ser técnico. Su vida no se puede cumplir sin *lo técnico*. Necesitamos la técnica para comer, para comunicarnos, para aprender... Incluso nuestro cuerpo ha ido co-

¹ El presente artículo constituye una versión avanzada y en español de las ideas presentadas por los autores en: "Technological changes and rights evolution in the bio-digital era", en Laura Liguori et al. (eds.), *Cambridge Handbook of Life Science, Information Technology and Human Rights*. Cambridge: CUP (en prensa).

² Università Campus Bio-Medico di Roma (m.bertolaso@unicampus.it).

³ Universidad de Valladolid (amarcos@fyl.uva.es).

evolucionando con nuestras técnicas. La estructura de nuestro aparato masticatorio, por ejemplo, está correlacionada con nuestro dominio del fuego y con nuestras capacidades para la fabricación y el uso de herramientas líticas. Un rechazo de lo técnico sería, por lo tanto, un rechazo de lo humano. No es adecuada una actitud radicalmente anti-técnica, pero sí es correcto preguntarse si la evolución de lo técnico favorece siempre la vida humana, si facilita la plena realización de las personas, o si puede resultar, en algunos casos, una amenaza para la misma. Para plantearnos esta cuestión es bueno observar cómo se ha ido desarrollando lo técnico, a partir de una primera modalidad, que está constituida por *las técnicas tradicionales*, a lo largo de la historia humana. La técnica, al unirse a la ciencia, a comienzos de la modernidad, dio lugar a *la tecnología*. La obra de Francis Bacon es, quizá, el lugar clásico para entender los fundamentos de esta confluencia entre técnica y ciencia. Tenemos ya aquí una segunda modalidad de lo técnico, que difiere en muchos aspectos de las técnicas tradicionales, que nos confiere más poder, mejores oportunidades y que nos expone, al mismo tiempo, a mayores riesgos. A finales de la modernidad la simbiosis entre ciencia y técnica se había hecho ya tan estrecha que empezamos a hablar de *tecnociencia*. Estamos ante la tercera modalidad de lo técnico. La tecnociencia incrementa aun más los poderes y los riesgos, y, por lo tanto, nos carga con una mayor responsabilidad moral.

Esta responsabilidad se hace aun más grave cuando comenzamos a modificar los seres vivos mediante las nuevas *biotecnologías*. Estas pueden ser consideradas como una cuarta modalidad de lo técnico, que afecta de un modo muy profundo a los procesos naturales. Hasta aquí, los productos de la técnica caían siempre fuera del reino de la vida, mientras que las biotecnologías lo que hacen es introducir artificialidad en algunos seres vivos.

Y crece aun más nuestra responsabilidad cuando el ser vivo que intentamos modificar técnicamente es el propio ser humano, es decir, el propio sujeto moral, el propio responsable de la acción tecnológica. Hablamos ahora de *antropotecnias*, es decir, de una quinta y más radical modalidad de lo técnico. Una bomba atómica es algo espectacular, ruidoso, devastador..., podría decirse que mucho más potente que una modesta modificación genética. Pero si esa modificación genética se aplica al ser humano, entonces estamos modificando al creador y potencial usuario de la propia bomba. Dicho de otro modo: las antropotecnias contemporáneas multiplican por un factor hasta ahora desconocido las oportunidades, riesgos y responsabilidades derivadas de las distintas modalidades de lo técnico.

Actualmente hay dos grandes propuestas antropotécnicas. Resumiendo y simplificando: tenemos el proyecto *bío* y el proyecto *info*. El primero trata de transformar al ser humano mediante las biotecnologías, el segundo se propone modificar, complementar, imitar o incluso sustituir al ser humano mediante las info-tecnologías. Con ello podemos decir que hemos ingresado en la era bio-digital, lo cual suscita nuevos retos éticos. ¿Cuáles son estos retos? Pretendemos, en este artículo, presentarlos y hacer una reflexión crítica sobre los mismos, así como algunas propuestas para abordar la era bio-digital sin pérdida de la perspectiva humanista.

Antes de adentrarnos en las cuestiones éticas que suscitan las nuevas antropotecnias, hemos de hacer aun dos observaciones. En primer lugar, el hecho de que existan dos grandes líneas de

intervención antropotécnica, *bío* e *info*, no es casual y responde ya a una cierta concepción antropológica que deberíamos someter a crítica. La disposición dual de las antropotecnias sugiere una antropología dualista. Habría que buscar los precedentes remotos de dicha antropología en el pensamiento platónico y gnóstico. Los precedentes modernos de la misma es probable que se encuentren en la filosofía cartesiana de las dos substancias, extensa y pensante, así como en las dos realidades básicas que identifica Kant, la naturaleza y la libertad. O bien, dicho en sus propias palabras, “el cielo estrellado sobre mi cabeza y la ley moral en mi corazón”. La filosofía antigua y medieval, en su vertiente platónica ha tendido al dualismo, no así en su vertiente aristotélica. También la filosofía moderna ha estado muy marcada por una antropología dualista. No es extraño que los proyectos antropotécnicos respondan aun a esta forma de pensamiento. Unos pretenden cambiar al ser humano modificando profundamente su cuerpo, ya sea a nivel genético o neuronal. Otros intentan poner el espíritu humano, concebido ahora como una especie de software, a salvo de las vicisitudes del cuerpo, liberarlo de la debilidad y finitud de la carne, digitalizarlo y subirlo a la nube, o bien sustituir completamente lo humano por algún otro género de entidad digital desencarnada, como, por ejemplo, un sistema de inteligencia artificial.

Podemos ya formular la segunda observación. Los problemas éticos de las nuevas antropotecnias derivan, en realidad, de otros problemas más profundos, de carácter metafísico. Es decir, para juzgar éticamente sobre las distintas propuestas antropotécnicas necesitamos saber previamente qué es un ser humano y en qué consiste el bien para el mismo. Las cuestiones éticas vienen después de las cuestiones metafísicas (ontológicas, epistemológicas y, sobre todo, antropológicas). A veces se piensa que una ética sin metafísica favorecería el consenso social. Pero lo cierto es que la búsqueda de una ética tal nunca ha culminado, y, sin embargo, la pérdida del horizonte metafísico no ha hecho sino incrementar el disenso moral y la falta de entendimiento entre nuestros conciudadanos. Dado que las propias tecnologías, y no digamos las antropotecnias, incorporan supuestos metafísicos más o menos implícitos, difícilmente podríamos abordar los retos morales que suscitan sin hacer previamente una crítica de los supuestos metafísicos a los que responden.

Veremos, pues, en lo sucesivo, cuáles son los retos éticos que presentan las bio-antropotecnias y, después, los que suscitan las info-antropotecnias. En ambos casos atenderemos también a las cuestiones previas de carácter metafísico, sin las cuales el abordaje de lo ético sería siempre precario.

2. Bio-antropotecnias

La palabra clave para las bio-antropotecnias es *enhancement* (mejora) (Bertolaso 2011). La intención, aquí, consiste en mejorar lo humano mediante la aplicación de diversas biotecnologías. Dichas tecnologías pueden afectar, en general, a cualquier parte de nuestro organismo, en forma, por ejemplo, de prótesis o de implantes. Ahora bien, el organismo humano, como el de cualquier otro animal, posee dos sistemas especialmente sensibles, a los que se dirigen la mayor parte de las propuestas de intervención tecnológica, nos referimos al sistema (epi)genético y al sistema nervioso. Sobre estos sistemas se pueden operar las

intervenciones más profundas, con el fin de implantar en nuestro organismo algún tipo de supuesta mejora.

Esta ilusión de mejora se ha visto inflamada en los últimos tiempos por las políticas de comunicación pública de los mega-proyectos de investigación. Conviene recordar que la llamada *big science* nació con el Proyecto Manhattan a mediados del siglo pasado, pero el relevo lo tomo, en la recta final de la centuria, el Proyecto Genoma Humano (PGH), cuyas conclusiones se publicaron en 2003. Y después les llegó el turno a los grandes proyectos de investigación sobre el cerebro humano (BRAIN y HBP). Los grandes proyectos científicos dependen críticamente de factores mediáticos, políticos y financieros. Parece que esto inclina a sus promotores a formular (o al menos a sugerir) promesas y expectativas deslumbrantes, como puede ser la de la mejora humana.

No obstante, esta intención de mejora puede ser interpretada al menos en dos sentidos diferentes. En primer lugar, hay quien piensa que el ser humano es meramente una parte de la naturaleza, que es un animal más. Como tal, podría ser mejorado en cuanto a sus rendimientos biológicos. Las intervenciones que se proyectan a partir de esta antropología naturalista radical pretenden convertir al ser humano en un animal mejor. Intentan tomar las riendas de la evolución biológica para obtener un tipo nuevo y mejor de organismos a partir de los individuos de nuestra especie. Más que mejorar al ser humano, pretenden transformarlo y superarlo, es decir, construir un ser post-humano ajeno ya a las limitaciones y dolencias propias de nuestra especie. En esta línea, el transhumanismo (TH) resulta la más influyente escuela de pensamiento.

En segundo lugar, podemos buscar a través de las biotecnologías, no la superación del ser humano, sino la *mejora de la vida humana*. Nosotros entendemos, junto con autores como Robert Spaemann (2012), que cada persona tiene ya un valor infinito al que llamamos dignidad (Marcos, 2018). En este sentido, no se puede mejorar, superar o añadir valor a ninguna persona en concreto mediante intervenciones biotecnológicas. Lo que sí se puede es mejorar la vida humana, hacer mejor la vida de las personas. Mejorar la vida humana significa hacerla más *propriadamente humana*. Es decir, la naturaleza humana tiene una función normativa. Así, una vida *propriadamente humana* será una vida conforme a la naturaleza humana. Si entendemos que el ser humano es un animal social y racional, entonces una vida *propriadamente humana* será una vida en la que estos tres aspectos se puedan desarrollar de modo armónico e integrado, una vida con salud física y mental, en pacífica y justa convivencia, con posibilidades de aprendizaje y de crecimiento espiritual, con una cierta libertad e identidad personal. Y a ello pueden contribuir, por supuesto, las aplicaciones de biotecnológicas. Y de hecho contribuyen, por ejemplo, cuando tienen una finalidad claramente terapéutica.

Obsérvese que estamos ante dos proyectos antropotécnicos distintos, con objetivos divergentes y bases filosóficas dispares. El primero se marca como meta la superación de lo humano. El segundo propone, en cambio, aplicar las tecnologías para profundizar en lo humano, para hacer de nuestra vida una vida más *propriadamente humana*. Pensamos que este último es el más adecuado y prometedor. Pero actualmente, y al margen del valor intrínseco de cada uno, el primero de estos proyectos es el que lleva la iniciativa mediática y académica, sobre todo a

través de la escuela transhumanista, mientras que el segundo está en una posición defensiva y crítica. Veamos, pues, en primer lugar las propuestas transhumanistas, para pasar después a la revisión de las críticas que estas reciben.

El TH se ha convertido en una auténtica moda cultural, en una ideología que marca la agenda filosófica (Asla, 2018). Incluso hay quien sostiene que la política futura se estructurará en función de la actitud de cada cual hacia las antropotecnias. Y el TH ya ha tomado posición en el debate: lidera las propuestas más favorables a las antropotecnias sin restricción.

Entre los pensadores transhumanistas destaca el núcleo de Oxford, y dentro del mismo las figuras de Julian Savulescu y de Nick Bostrom (Savulescu y Bostrom, 2009). Por su parte, en Estados Unidos, Max More fundó en 1990 el *Instituto Extropiano*, dedicado a la promoción de las ideas transhumanistas. A More (2010) le debemos esta definición: “El transhumanismo es una clase de filosofías que buscan guiarnos hacia una condición posthumana”. Reconocemos aquí un claro afán de superación de la naturaleza humana, entendida siempre en términos de constricciones negativas, nunca de limitaciones constitutivas. Existe, además, otro frente de defensa de la antropotecnia profunda, formado por los filósofos posthumanistas continentales, para quienes el ser humano es un animal aporético, entre los que podemos citar, por ejemplo, a Peter Sloterdijk (2003) o a Giorgio Agamben (2003), y cuya inspiración hay que buscarla en la tradición existencialista y nihilista.

Los aspectos en los que se busca la supuesta mejora del ser humano abarcan desde las capacidades físicas, morales e intelectuales, hasta las emociones, la longevidad y prolongación de la juventud (Bertolaso, 2011). Los procedimientos pueden ir, desde la modificación de la dieta, hasta la transformación de estructuras moleculares u orgánicas, cambios permanentes de bases genéticas, empleo de diversas sustancias químicas, desarrollo de prótesis motoras o sensoriales, externas o internas, en forma de *hardware* o de *software*, potenciación del crecimiento de la corteza cerebral, modificación del sueño, etc. Se podría intentar, por ejemplo, el “mejoramiento” de las capacidades sensoriales, de modo que pudiéramos ver en la oscuridad y también en la franja del infrarrojo o del ultravioleta, así como oír fuera de las frecuencias habituales.

El TH ha sido criticado desde muchos frentes. Hay quien piensa que estamos ante meras ensoñaciones futuristas sin viabilidad. Existe también la posibilidad de impugnar el TH desde el concepto de naturaleza humana (Kass 2002, Sandel 2007). Una práctica tecnológica que pretende ir más allá de la naturaleza humana destruye de paso los criterios de valoración. En consecuencia, difícilmente se puede llamar mejora. Se trataría, en principio, de un cambio sin componente axiológico. Y, en el fondo, de un cambio a peor, pues la pérdida del componente axiológico, la *indiferencia*, supone ya una pérdida de valor. En palabras de Hans Jonas (1995, 95-6), la mera posibilidad de valor es ya un valor. Parafraseando a Jonas, podríamos decir ahora que el tránsito hacia la imposibilidad de valor es ya un disvalor.

Hay objeciones que vienen desde el ángulo epistemológico y ecológico, ya que la propia naturaleza de las ciencias complejas (como la biomedicina o la ecología, las ciencias climáticas o la epidemiología) no permite predecir los posibles efectos y riesgos de manera determinista

y mecanicista (Bertolaso *et al*, 2015). La incertidumbre y la *realizabilidad múltiple* (es decir, la posibilidad de que los sistemas vivos encuentren diferentes formas de alcanzar sus objetivos de supervivencia, reproducción y reparación en los diferentes niveles de la organización biológica, organísmica o ecosistémica) son muy comunes, lo que limita la posibilidad de controlar los sistemas vivos como hacemos con las máquinas. Es fácil dar actualmente con ejemplos de todo ello en los debates sobre la pandemia de COVID-19 y sobre la posibilidad real de encontrar una solución única y definitiva en una vacuna.

También hay dudas acerca de la viabilidad ecológica y social de una longevidad humana muy superior a la actualmente conocida. ¿Qué pasaría si cada persona viviese trescientos, cuatrocientos... años? Hay quien sostiene que, en virtud del principio de precaución, debería al menos producirse una moratoria respecto de algunos proyectos antropotécnicos. Por otra parte, y desde el pensamiento social, se puede dudar de la justicia e igualdad con que se realizarían las presuntas mejoras. En lo político, tanto Jürgen Habermas (2002), como Francis Fukuyama (2002) insisten sobre los peligros que tiene la mentalidad transhumanista para la libertad y la democracia. En el terreno de la bioética, han aparecido también críticas al proyecto de presunta mejora humana (Postigo, 2019). El principio bioético de autonomía queda en entredicho (Habermas, 2002). En cuanto a los principios de beneficencia y de no maleficencia, es más que dudoso que se puedan aceptar experimentos inciertos sobre seres humanos sin ninguna ganancia clínica. El principio de justicia también quedaría comprometido en la medida en que los recursos disponibles para la presunta mejora compitiesen con los que se dedican a las acciones propiamente terapéuticas.

Como es fácil ver, todos estos desafíos tienen implicaciones al mismo tiempo tecnocientíficas y antropológicas. Su buena resolución depende de que adoptemos una antropología correcta, alejada del dualismo y del naturalismo radical. Solo así podremos obtener lo mejor de las nuevas tecnologías: su capacidad para favorecer la vida humana y el cuidado mutuo.

3. Info-antropotecnias

En esta sección veremos cómo el despliegue de sistemas digitales está planteando preguntas nuevas y desafiantes para nuestra comprensión del ser humano y de las dinámicas sociales (Marcos 2020). En especial, el uso de algoritmos de aprendizaje automático y de aprendizaje profundo (*deep learning*), asociados con la minería y procesamiento de grandes cantidades de datos (*big data*), plantea, en los ámbitos de la salud y de la educación, importantes desafíos para garantizar valores como la igualdad, la libertad, la privacidad o la intimidad.

La relación de las tecnologías *info* con la vida humana es compleja. A veces se presentan simplemente como herramientas para mejorar dicha vida, como por ejemplo en el caso de la telemedicina. Pero en otras ocasiones se les atribuyen pretensiones más audaces, como la de simular las capacidades y funciones humanas, superar al ser humano o, incluso, llegar a sustituirlo. Estas expectativas se aplican especialmente a la llamada inteligencia artificial (IA).

El término “inteligencia artificial” sirve para referirnos a la supuesta inteligencia que tendrían algunos artefactos o sistemas artificiales. Como es bien sabido, el término IA fue acuñado en 1956, durante la ya famosa conferencia de Dartmouth, por el científico estadounidense John

McCarthy. Es cierto que, desde el punto de vista retórico y propagandístico, este término ha resultado todo un éxito. Pero la inteligencia es la capacidad de entender, y sabemos que un sistema artificial es incapaz de entender nada. Ni siquiera se puede decir con propiedad que una máquina cuente o compute. Contar implica reunir dos (o más) momentos diferentes, y mantenerlos como tales, en una sola representación consciente, cosa que una máquina no hace.

Sí es cierto, en cambio, que la llamada IA puede ayudarnos a resolver diversos problemas (cómputo, geolocalización, logística, asistencia telefónica, asistencia al diagnóstico médico, a la publicidad y un largo etcétera). Pero estos problemas no son tales para el sistema artificial, sino para el diseñador o para el usuario humano del mismo. Para un sistema de reconocimiento facial, el reconocer o no a un delincuente no supone un problema. Es un problema para la seguridad de las personas, y el sistema puede ayudarnos a afrontarlo. Por supuesto, el mismo sistema puede servir para controlar a la población de un país y para facilitar la represión política. Pero esto tampoco es un problema para las cámaras o para el software implicado. Lo es, indudablemente, para los sufridos súbditos del país en cuestión. Ni los martillos, ni los ábacos, ni los sistemas informáticos más avanzados tienen problemas. Los problemas los tenemos nosotros, como seres vivos y como seres humanos. Solo un ser vivo puede morir o sufrir, solo una persona puede preguntarse por el sentido de su vida. Estos son problemas. Y tanto un martillo como una red informática, cada uno a su modo, pueden ayudarnos a afrontarlos (o a empeorarlos). Pero esto no los convierte en inteligentes.

Podemos verlo también si atacamos la cuestión desde otro ángulo. A veces se caracteriza la llamada IA por su capacidad de simulación. Simula funciones propias de la inteligencia humana, se dice. Pero simular no es *ser*. Simular la inteligencia no es lo mismo que ser inteligente. Y, además, la noción de función remite inexorablemente a la de un ser para el cual un efecto dado es funcional. Aquí, los sistemas artificiales dependen también de la funcionalidad que puedan tener para el ser humano. Fuera del marco humano, las luces que se encienden y apagan en una pantalla o los movimientos de un robot son meros efectos, no cumplen funciones. Es más, el concepto de *dato*, como el de mensaje, implica el de un receptor de ese dato, y dicho receptor, directa o indirectamente, será un ser humano. Es mi mirada sobre mi ordenador la que transforma en datos para él todos los *inputs* que le llegan, y en datos para mí los *outputs* que emite. Solo suponiendo una conciencia en el trasfondo, podemos hablar de ciertos estados de la materia como datos, y de un sistema de IA como un receptor de datos.

La cuestión, en el fondo, es de carácter ontológico. Los artefactos, en la tradición aristotélica, son sustancias solo en un sentido accidental, por analogía. Los seres vivos, y en especial los seres humanos, es decir, las personas, sí son sustancias en sentido propio y paradigmático. Al tratarse de una diferencia ontológica, la esperanza (o la amenaza) de anularla por la vía de la sofisticación tecnológica resulta ilusoria, un mero error categorial.

Estrictamente hablando, como afirma Luc Julia (2019), uno de los creadores del sistema de IA denominado SIRI, la inteligencia artificial no existe. Pero, ¿cómo se ha llegado a la *pretensión* de que existen sistemas artificiales inteligentes? Con el auge de los ordenadores personales, la miniaturización y abaratamiento de los componentes y la llegada de Internet, se dispararon en pocos años tanto la capacidad de computación, como el aporte de datos. Algunas empresas

vieron pronto en este despegue una promesa de abundancia, y su propia actividad generó aun muchos más datos y mayor capacidad para procesarlos. Google, Amazon, Facebook, Apple y Microsoft (conocidas en conjunto como GAFAM), así como las chinas Baidu, Alibaba, Tencent, Xiaomi (BATX) y Huawei, han encendido de nuevo la ilusión por la IA. Lo han hecho a base de fuerza bruta, es decir, a costa del *big data* y de una capacidad de computación exponencialmente creciente y cada día más asequible.

Gracias a este repunte de la investigación en IA se han logrado sistemas muy útiles y precisos en diversos ámbitos. Siempre se trata de sistemas especializados en una cierta función. Han mejorado, indudablemente, los sistemas de traducción automática, los de reconocimiento facial o visual, los sistemas expertos en medicina o en finanzas, los sistemas de conducción automática o los robots que simulan habilidades conversacionales (*chatbots*). Ninguno de ellos entiende lo que hace, pero eso no disminuye su funcionalidad... para nosotros. No se ha logrado ningún sistema de IA fuerte o general, es decir, un sistema artificial que pudiera afrontar de modo integrado todas las funciones que realiza la inteligencia humana.

Pensemos, para hacernos una idea más concreta, en un sistema de IA útil, por ejemplo, para la banca. Nos ayuda a decidir si debemos o no conceder un determinado crédito. Pero, ¿quién o qué toma realmente la decisión? ¿La tomará el sistema de IA por nosotros? ¿Está de más, entonces, el asesor financiero humano?, ¿sobra la opinión del médico cuando un sistema de IA diagnostica un cáncer?, ¿huelga el piloto humano cuando el sistema de navegación inteligente decide variar la velocidad o la altura de vuelo?, ¿podría calificarse siempre de temerario el obrar contra el *output* de un sistema de IA?, ¿qué o quién se hace responsable en cada caso de un posible fallo?

Muchos sostienen que la transparencia algorítmica ha de convertirse en la piedra angular de la regulación de la Inteligencia Artificial. Según Rotenberg (en prensa): “La transparencia algorítmica es el objetivo más urgente en cuanto a la regulación de la IA, se convertirá en un indicador crítico de la salud de las sociedades democráticas”. Pero estos buenos deseos plantean nuevos interrogantes, dado que los requisitos de auditabilidad y transparencia pueden chocar con el difícil (¿insuperable?) problema del tiempo y que los sistemas de IA se utilizan precisamente para ahorrarnos tiempo: ¿podremos nosotros, los humanos, encontrar el tiempo necesario para auditar las tareas que tales sistemas realizan a una velocidad vertiginosa, o terminaremos encomendando a otros sistemas de IA las labores de auditoría y supervisión?

No hace falta llamar mucho la atención sobre las importantes consecuencias de todo tipo que tienen estas preguntas. Vislumbramos inmediatamente sus implicaciones filosóficas, antropológicas, sociales, jurídicas, económicas, morales, laborales... (David y Sauviat, 2019). Los riesgos, de hecho, irán más allá de cualquier medida o escenario manejable si descuidamos el desarrollo de un marco filosófico y antropológico sensato, que nos dé una comprensión adecuada de los procesos de decisión (Bertolaso y Sterpetti, 2020).

La llamada IA funciona bien como complemento de la inteligencia humana. Cobra entidad y sentido solo en el marco de la acción humana. La primera razón, y la más obvia, es que se originó como producto de la acción humana. Pero hay quien piensa que una vez puesta en

marcha podría auto-mantenerse, incluso auto-mejorarse, al margen ya de lo humano. No obstante, el problema en cuanto a los límites de la IA no es técnico, es ontológico.

Lo cierto es que ni siquiera existirían propiamente datos (ni *big data*) en ausencia de un receptor humano, de una persona que sea capaz de unificar en la conciencia la dispersión de los hechos brutos, que sea capaz de darle intencionalidad a cada uno de ellos, de conectar así unas partes de la realidad, que juegan como datos, con otras acerca de las cuales nos dicen algo estos datos. Además, por más laxos que podamos ser en la selección de datos para el histórico, siempre habrá alguna selección, pues no podemos alimentar una máquina con *todos* los datos. En la recogida de los mismos y en su aportación a los sistemas de IA realizamos continuamente juicios de relevancia inexorablemente humanos.

A partir de estas consideraciones ontológicas podemos derivar una idea que es crucial desde el punto de vista ético: solo las personas pueden tomar decisiones. El mismo concepto de decisión es ajeno a lo mecánico. Lo que llamamos “decisión” en un sistema de IA lo será solo en la medida en que un ser humano haya tomado la genuina decisión de delegar alguna acción en el sistema, es decir, de automatizarla. La responsabilidad última, salgan bien o mal las cosas, solo puede ser de un ser humano. Y es responsabilidad nuestra apoyarnos en los mejores sistemas de IA disponibles a la hora de tomar decisiones. Es decir, en muchos casos, será conveniente que el médico se fie de un sistema de IA para el apoyo al diagnóstico. Especialmente si el sistema está bien calibrado y se ha mostrado útil y fiable en ensayos clínicos. Pero, de cualquier modo, el experto humano debe tener las manos libres para obrar incluso contra las indicaciones de un sistema de IA. Muchas veces hace falta *alguien* que entienda, que busque creativamente nuevas explicaciones, que replantee y genere nuevos sistemas de expectativas.

4. Conclusiones y perspectivas: hacia una concepción humanista de la tecnología

Hoy más que nunca necesitamos una reflexión filosófica sobre nuestra relación con la tecnología. Las actitudes extremas no nos sirven. La actitud anti-tecnológica no hace justicia al hecho de que el ser humano y la técnica han co-evolucionado. Pero la actitud tecnologista tiende a la deshumanización. Como hemos tratado de mostrar, la reflexión filosófica a la que nos referimos ha de incluir aspectos éticos, pero también y fundamentalmente aspectos ontológicos, epistemológicos y antropológicos (Marcos, 2018a, 2020a). Por ejemplo, hemos visto que no puede existir una ontología de lo digital -más allá de una crudamente física- si omitimos el factor humano. El punto de vista del ser humano es el que convierte meros hechos físicos en datos intencionales, en procesos de computación, en información y en conocimiento, es el que transforma el bullir de los electrones en memoria e inteligencia. En lo epistemológico, nos hemos percatado de que la IA está afectada por el clásico problema de la inducción y limitada en cuanto a su propia capacidad intelectual, pues las máquinas, en realidad, no *entienden*. En lo antropológico resulta crucial la superación de una antropología dualista y el desarrollo de una visión integrada del ser humano como persona.

Todo ello configura la noción emergente de *humanismo tecnológico*, desde la cual se pueden afrontar los problemas éticos y sociopolíticos que suscitan las nuevas tecnologías, así como aprovechar las aportaciones que hacen a una auténtica mejora de la vida humana. Las

preocupaciones éticas vienen después de las cuestiones metafísicas. Dado que las tecnologías en sí mismas –y sobre todo las antropotecnias- incorporan supuestos metafísicos más o menos implícitos, difícilmente podríamos abordar los desafíos morales que plantean sin antes criticar los supuestos metafísicos a los que responden.

El humanismo tecnológico nos invita a hacer uso de la tecnología para garantizar aquellas circunstancias que realmente mejoren la vida humana, tanto individual como colectiva, y que protejan decididamente los derechos y libertades. Tal noción de humanismo tecnológico está condicionada por una cierta comprensión específica de las relaciones humanas y de las relaciones entre el ser humano y los artefactos tecnológicos. Obviamente las relaciones entre las personas pueden ser mediadas por la tecnología, pero esta no puede reemplazar ontológicamente a las propias personas. Así, mejorar la vida humana para hacerla más propiamente humana requiere una nueva consideración de las dimensiones relacionales de lo humano, y muy especialmente de las que llamamos *cuidado y trabajo* humano (Bertolaso y Rocchi, 2020), implica, además un cuidadoso equilibrio entre las relaciones de dependencia mutua y la legítima aspiración a la autonomía personal.

En continuidad con lo dicho, las transiciones bio-digitales actuales nos están obligando a profundizar en los fundamentos de los derechos humanos para dar una mejor explicación de lo que realmente significa ser humano hoy en día en medio de una realidad socio-tecnológica. Es evidente que las realidades tecnológicas nos obligarán a repensar los derechos humanos y a ampliarlos mediante alguno de estos dos mecanismos: o bien la profundización en la potencialidad de los derechos humanos de primera generación, o bien la creación de una cuarta generación de derechos humanos específicamente referidos a las tecnologías bio-digitales. No podemos abordar aquí el debate por extenso, pero sí, al menos dejar expuesta nuestra opinión al respecto. Pensamos que la proliferación de derechos humanos, distribuidos en diversas generaciones, no hace sino debilitar la propia noción de derecho humano, así como su fuerza y autoridad (Puppink, 2020). Siempre sería mejor profundizar en la comprensión de los derechos más básicos, los de primera generación, y en especial el derecho a la vida y a la libertad, y tratar de aplicarlos a las nuevas condiciones de vida en la era bio-digital. Esta posición será siempre más afín a las tradiciones iusnaturalistas, mientras que la proliferación de derechos coincide mejor con el espíritu del iuspositivismo. Los llamados derechos humanos han constituido desde hace décadas un pilar clave de nuestras sociedades y culturas. A nuestro modo de ver, el desafío ahora consiste en fortalecer sus raíces y no enfatizar la multiplicación de "derechos".

Desde la perspectiva del humanismo tecnológico que aquí proponemos, es incluso posible que las nuevas tecnologías nos ayuden a apreciar y manejar mejor ciertos aspectos de nuestra naturaleza humana, que antes se tenían por secundarios o no propiamente humanos. Piénsese, por ejemplo, en nuestra vulnerabilidad, en nuestra dependencia mutua, en la importancia de la solidaridad o de la atención cara a cara como estándar de oro para el cuidado médico o parental.

6. Referencias bibliográficas

Agamben, G., 2003, *Homo sacer*, Pretextos, Valencia.

- Asla, M., 2018, “El transhumanismo como ideología: ambigüedades y dificultades de la fe en el progreso.” *Scio*, 15: 63-96.
- Bertolaso, M., 2011, “Le Human Enhancement Technologies e l’Irriducibilità della Complessità Biologica”, en S. Kampowski, S., Moltisanti D. (eds.), *Migliorare l’uomo? La sfida etica dell’enhancement*, Cantagalli, Siena, pp. 35-58.
- Bertolaso, M. et al., 2015, “Bio-Techno-Logos and Scientific Practice”, en Bertolaso, M. (ed.), *The Future of Scientific Practice*, Pickering & Chatto Publishers, Londres, pp.179-191.
- Bertolaso, M. y Rocchi, M., 2020, “Specifically Human: Human Work and Care in the Age of Machines”, *Business Ethics: A European Review* (in press).
- Bertolaso, M. y Sterpetti, F. (eds.), 2020, *A Critical Reflection on Automated Science – Will Science Remain Human?*, Springer, Dordrecht.
- David, M. y Sauviat, C., 2019, *Intelligence Artificielle*, Éditions du Rocher, Mónaco.
- Fukuyama, F., 2002, *El fin del hombre: consecuencias de la revolución biotecnológica*. Ediciones B, Barcelona.
- Habermas, J., 2002, *El futuro de la naturaleza humana. ¿Hacia una eugenesia liberal?*, Paidós, Barcelona.
- Jonas, H., 1995, *El principio de responsabilidad. Ensayo de una ética para la civilización tecnológica*. Herder, Barcelona.
- Julia, L., 2019, *L’intelligence artificielle n’existe pas*, Éditions First, París.
- Kass, L., 2002, *Life, Liberty and the Defense of Dignity: The Challenge for Bioethics*, Encounter Books, San Francisco.
- Marcos, A., 2018, “¿Tienen igual dignidad todos los seres humanos?”, en Pérez de Laborda, M., Vanney, C. y Soler, F. J. (eds.), *¿Quiénes somos? Cuestiones en torno al ser humano*. Eunsa, Pamplona, pp. 210-214.
- Marcos, A., 2018a, “Bases filosóficas para una crítica al transhumanismo”, *Artefactos. Revista de Estudios de la Ciencia y la Tecnología*, 7: 107-125.
- Marcos, A., 2020, “Información e Inteligencia Artificial”, *Apeiron. Estudios de Filosofía*, 12: 73- 82.
- Marcos, A., 2020a, “New Praxis. Releasement in a Technological World”, en Valera, L. y Castilla, J. C. (eds.), *Global Changes. Ethics, Politics and Environment in the Contemporary technological World*. Springer, Dordrecht, pp. 17-27.
- More, M., 2010, *Transhumanist FAQ*, URL=<http://www.extropy.org/resources.htm/>
- Postigo, E., 2019, “Bioética y transhumanismo desde la perspectiva de la naturaleza humana”, *Arbor*, 195 (792): a507. <https://doi.org/10.3989/arbor.2019.792n2008>.

Puppinck, G., 2020, *Mi deseo es la ley*, Encuentro, Madrid.

Sandel, M., 2007, *Contra la perfección. La ética en la era de la ingeniería genética*, Marbot, Barcelona.

Savulescu, J., Bostrom, N., 2009, *Human Enhancement*. OUP, Oxford.

Sloterdijk, P., 2003, *Normas para el parque humano*, Siruela, Madrid.

Spaemann, R., 2012, *Love and the Dignity of Human Life*, Eerdmans, Grand Rapids, MI.